

Nevromorft kamera

Jens Petter Abrahamsen

Institutt for Informatikk, Universitetet i Oslo

Hovedoppgave levert som del av studiet til cand.scient.

2001 – 2003

Veiledere: Tor Sverre Lande og Philipp Häfliger.

FORORD

Hovedoppgaven *Nevromorft kamera* har vært en utfordring. Uten bakgrunn i elektronikk, har det vært mye å sette seg inn i, men det har uten tvil vært interessant. Den store friheten jeg har hatt til å utforme oppgaven akkurat slik jeg ønsker, har til tider vært frustrerende og også gjort at det har tatt lenger tid. Det positive er at arbeidet med oppgaven har blitt tilsvarende mer spennende og forskningsrettet.

Jeg benytter anledningen til å takke kjæresten min Qibei, for at hun har støttet opp og passet vår datter Heidi Louise ekstra når det har vært frister som måtte overholdes, og for alle de gode matpakkene. En stor takk også til veiledere Bassen og Philipp for hjelp og ideer de har kommet med.

INNHALDSFORTEGNELSE

Forord	ii
Innholdsfortegnelse	iii
1 Innledning	1
1.1 Analoge integrerte kretser	1
1.2 Nevromorfe systemer	1
1.2.1 Likheter mellom mikroelektronikk og biologien	2
1.2.2 Biologien er en stor suksess.....	2
1.3 Optiske kretser.....	3
1.3.1 Synet - biologiens optiske krets.....	3
1.3.2 Mer detaljert om syn.....	3
1.3.3 Periferisynet	5
1.4 Problemstilling	5
1.5 Et større tenkt system	6
1.6 Praktiske anvendelser.....	6
2 Generelt bakgrunnsstoff.....	7
2.1 Strømmen gjennom transistoren.....	7
2.2 Strømmen i biologiske nevroner	8
2.3 Forskjeller og likheter.....	9
2.4 Teknologivalg	9
2.4.1 Fotobrikker i CMOS.....	9
2.4.2 CCD og kombinasjonen CMOS/CCD	10
2.4.3 Andre prosesser: BiCMOS og GaAs.....	10
2.5 Smarte sensorer.....	10
2.6 Romlig prosessering.....	11
2.7 Temporær prosessering	11
3 Hele systemet	13
3.1 Inndeling av systemet i tre deler.....	13
3.1.1 Optikk i øyet og kameraet	13
3.1.2 Linseholder (kamerahus).....	14
3.1.3 Netthinnen og sensoren.....	14
3.1.4 Ytterligere prosessering.....	14
3.2 Utforming av hele systemet	15
3.2.1 Et eller flere objekter.....	15
3.2.2 Ekte bilde eller bare ”meta” informasjon.....	16
3.2.3 Beregning av sentrum.....	17
3.2.4 Utlesningsmekanisme.....	17
3.2.5 Nevromorf koding av utgangssignalene	18
3.3 Oversikt over hovedkretsen.....	19
3.4 Dedikerte testkretser	20
4 Lysfølsomt element og fotokretsen.....	21
4.1 Lysfølsomme elementer/strukturer i CMOS	21

4.1.1	Fotodetektorer.....	21
4.1.2	Utfordringer for fotodetektorer.....	22
4.2	Utforming av lysfølsomt element.....	22
4.3	Utlegg av fotodioden.....	24
4.4	Fotokretser.....	24
4.4.1	Sample-and-hold fotokretser.....	25
4.5	Tilpasning til lyset.....	25
4.5.1	Global adaptasjon.....	26
4.5.2	Lokal adaptasjon.....	26
4.5.3	Aktive sensorer.....	27
4.6	Fotokretsen i kameraet.....	27
4.6.1	Analyse av fotokretsen.....	28
4.6.2	Simuleringer av fotokretsen.....	29
4.7	Utlegg av fotokrets.....	30
4.8	Målinger på fotokretsen.....	31
4.8.1	Sammenlikning med simuleringene.....	32
4.8.2	Kort analyse av grafen.....	33
4.8.3	Ulik tilpasningshastighet opp og ned.....	34
4.9	Om testing med lys.....	35
5	Korrelasjonskrets og endringsdetektor.....	36
5.1	Deteksjon av endringer.....	36
5.2	Beskrivelse av korrelasjonskretsen.....	36
5.3	Utlegg av korrelasjonskretsen.....	38
5.4	Målinger av korrelasjonskretsen.....	39
5.4.1	Resultater.....	39
5.5	Konstruksjon av endringsdetektoren.....	41
5.6	Simuleringer.....	42
5.7	Utlegg.....	43
5.8	Målinger av endringspiksel (m. fotodiode).....	44
5.8.1	Sammenkopling av flere endringspikslar i en rad/kolonne.....	46
5.8.2	Innstilling av rad- og kolonnestrømmene.....	46
6	Nevroner og winner-take-all kretsen.....	49
6.1	Nevromorf koding.....	49
6.2	Integrate-and-fire nevronet.....	50
6.3	Winner-take-all kretser.....	51
6.4	Tidsdomene WTA av integrate-and-fire nevroner.....	52
6.4.1	Et nevron i winner-take-all kretsen.....	53
6.4.2	Oppløsning og skalering.....	54
6.4.3	Nullstilling av nevronene.....	55
6.4.4	Nødvendigheten av todelt nullstilling.....	55
6.4.5	Inaktivitets (timeout) nevronet.....	56
6.5	Digital støy fra nevronene.....	57
6.6	Utlegg.....	58
6.7	Målinger av winner-take-all kretsen.....	59
6.7.1	Glidende overgang.....	60
6.7.2	Eksempler på ufullstendig nullstilling.....	61
6.7.3	Måling av forsinkelsen.....	65

6.7.4	Forsøk på å få flere vinnere ved å øke strømmen.....	65
7	AER protokollen og målinger av hele systemet.....	67
7.1	Utgangssignaler fra systemet.....	67
7.2	Adresse-hendelse (AER) protokollen.....	67
7.2.1	Kollisjonsstrategier	68
7.3	AER som kommunikasjonsprotokoll for systemet	69
7.4	Målinger på hele kretsen.....	70
7.4.1	Virkningen av den uønskede spenningsforskjellen $V_{out} - V_{feedback}$..	71
7.4.2	Generelt om resultatene.....	71
7.4.3	Forsøk som ble gjort	71
7.4.4	Rad og kolonne par.....	72
7.4.5	Forsøk på å forbedre sirkel- og firkant resultat	73
7.4.6	Bevegelse langs radene eller kolonnene	73
7.4.7	Problem med kollisjoner?	74
8	Oppsummering og videre arbeid	80
8.1	Kommentar av resultatene	80
8.2	Hva har arbeidet bestått i	80
8.3	Utvidelser / videre arbeid.....	80
8.3.1	Flere objekters følger.....	80
8.3.2	Winner-take-all med hysteresis	80
8.3.3	Rad og kolonne annenhver gang	80
9	Litteratur.....	82
10	Vedlegg.....	83
10.1	Foto av kretsen	83
10.2	Linsen og linseholderen	84
10.3	Målepunkter, testkretser og innstilling av kretsen	86
10.3.1	Målepunkter i hovedkretsen	86
10.3.2	Målepunkter i spesielle testkretser	87
10.3.3	Prosedyre for innstilling av kretsen.....	88
10.4	A time-domain winner-take-all circuit of integrate-and-fire neurons	89

1 INNLEDNING

1.1 Analoge integrerte kretser

I analog kretskonstruksjon ønsker man å utnytte karakteristikkene til elementene i standard prosesser. I motsetning til digital kretskonstruksjon, der man først og fremst bruker transistoren som en bryter som kan være av eller på, ser man på overføringsfunksjonene til transistoren og andre elementer som dioden. En annen viktig forskjell fra digitale kretser, som er absolutt nøyaktige, forutsigbare og enkelt programmerbare, er at analoge kretser har et snev av unøyaktighet og uforutsigbarhet.

Absolutt nøyaktighet og forutsigbarhet er nødvendig for de fleste algoritmer. Når det likevel ikke kreves, finnes det utrolige muligheter i de analoge elementene for den som kan utnytte dem. For eksempel kan man regne ut logaritmer, eksponenter og tangens hyperbolicus med en enkelt transistor. Med noen flere transistorer kan man legge sammen og trekke fra, gange og dele. Haken er at verken inngangverdiene eller svaret er helt nøyaktig, men til mange formål er det godt nok. Tilsvarende digitale kretser trenger hundrevis av (noe mindre) transistorer.

Når man utvikler fotobrikker, er det lurt å ta hensyn til at de lysfølsomme elementene, som i likhet med mange sensorer, i utgangspunktet er analoge og har store interne forskjeller. Det betyr at absolutt nøyaktighet er vanskelig å oppnå. Digitalisering av signaler tilfører også ekstra unøyaktigheter, såkalt kvantiseringsskøy. Analoge beregninger direkte på brikken har særlig fordel av at massive parallelle beregninger kan gjøres.

Analog kretskonstruksjon kan, i motsetning til digital, vanskelig automatiseres. Ferdige kretser og utprøvde teknikker er til stor hjelp, men analoge kretser er likevel svært krevende å lage. Man må vite å forenkle på de rette stedene og lage stabile kretser som tåler de store unøyaktighetene i moderne prosesser. Å tenke ut slike løsninger er svært krevende. Det hjelper hvis man har noe å se etter. En rekke "systemer" som fungerer glimrende er biologien, som er utviklet gjennom millioner av år med naturlig utvalg. En nytenkning innenfor analog kretskonstruksjon er å etterlikne løsninger derfra.

1.2 Nevromorfe systemer

Innenfor gruppen "nevromorfe systemer" arbeider vi med å etterlikne løsninger som finnes i biologien, spesielt i hjernen. I denne oppgaven etterliknes et grunnleggende element ved syn nemlig evnen til å føle bevegelser og merke seg hvor de er. Det er et mål å gjøre etterlikningen så gjennomført som mulig, uten å gjøre ting unødvendig komplisert. I naturen begynte det med de lavere deler av hjernen, så da er det også et godt utgangspunkt for vår forskning.

Det kan virke søkt å etterlikne biologien. Ved første øyekast virker mikroelektronikk, som er konstruert av mennesker, nøyaktig og strukturert, totalt forskjellig fra biologien, der det meste er unøyaktig og til tider basert på følelser og

instinkt. Men, som vi skal se, er det også mange likheter mellom biologien og analog kretskonstruksjon, og disse blir stadig mer aktuelle.

1.2.1 Likheter mellom mikroelektronikk og biologien

Det er flere likheter mellom analog kretskonstruksjon i mikroelektronikk og nervesystemet. Den mest åpenbare likheten er unøyaktigheten. Akkurat som to transistorer sjelden er helt like, finnes det heller ikke helt faste mål på et biologisk nevron, som er en spesialcelle i nervesystemet. Etter hvert som prosessene utvikler seg blir den relative unøyaktigheten ikke mindre, fordi utviklingen styres av digitale kretser.

En annen likhet er at nervesystemet også er et (kjemisk-) elektrisk system. Kjente begreper som strøm, spenning, ladning og isolatorer, kan brukes i beskrivelsen av nervesystemet. Strømmen er riktignok negativt ladede ioner istedenfor elektroner, men virkemåten tilsvarende. For eksempel lader en strøm av ioner opp en spenning over celleveggen til nevronene. Det finnes ledninger, akkurat som i mikroelektronikk, og disse brukes til å overføre signaler. Signaler som overføres over lengre avstander blir forsterket og repetert av spesielle ”kretser” som finnes i nervesystemet.

Et annet fellestrekk mellom biologien og mikroelektronikk er at både transistoren og nevronet baserer seg på de samme to fysiske fenomenene: drift og diffusjon. Drift er den stadige bevegelsen av partikler i et medium som følge av en utvendig kraft, for eksempel et elektrisk felt. Diffusjon er kjent fra læren om gasser. Den stadige bevegelsen av partikler vil gjøre at det går flere partikler fra et område med høy konsentrasjon til lav konsentrasjon enn omvendt og fører derfor til utjevning av konsentrasjoner. Diffusjon og drift i transistoren og nevronet er relatert til hverandre ved Einsteins relasjon.

Siden nevronene i nervesystemet og transistorene i mikroelektronikk baserer seg på de samme prinsippene, er det mulig å sammenlikne, og eventuelt etterlikne, løsninger i de to domenene.

To viktige karakteristikk som deles av nevroner og transistorer er de store unøyaktighetene og at det er et eksponentielt forhold mellom strøm og spenning. Sistnevnte gjelder for transistorer i svak inversjon. I sterk inversjon har transistoren et kvadratisk strøm/spenning forhold, og mange av prinsippene fra biologien blir da vanskelig å overføre.

1.2.2 Biologien er en stor suksess

Det er liten tvil om at biologien er en suksess. Etter millioner av år med utvikling er biologiske skapninger godt tilpasset til miljøet de lever i. For å nevne noen konkrete eksempler på hvor godt biologien fungerer sammenliknet med en moderne mikroprosessor, kan insekter nevnes. De fleste insekter har hjerner som er veldig små sammenliknet med en mikroprosessor. Likevel kan de utføre kompliserte oppgaver som å fly uten å kollidere, lande, unngå farer, de kan lukte, finne tilbake til f.eks. bikuben og mer. Bier og vepser har også sosiale ferdigheter. Og effektforbruket er en brøkdel av det moderne prosessorer bruker. Mikroprosessorer er imidlertid bedre egnet til å utføre nøyaktige utregninger.

Alt dette til tross for at biologisk materiale er veldig unøyaktig, uten gode ledere. Det må derfor være løsninger her som er meget effektive, og disse er derfor fristende å etterlikne. Det er da særlig hjernen vi studerer. For utvikling av sensorer kan man hente idéer fra sansene til mennesker og dyr.

En av de generelle løsningene i biologien er parallelle beregninger på samme datasett. Selv om hver beregningsvei er unøyaktig, blir resultatet mer nøyaktig enn en enkelt beregningsvei.

1.3 Optiske kretser

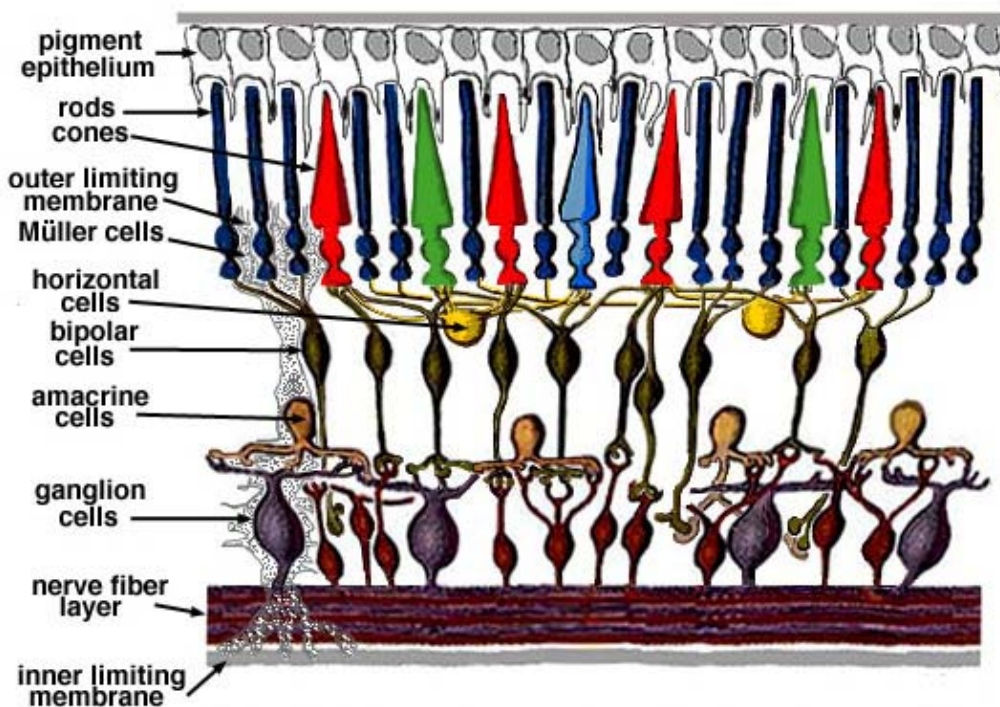
Etter hvert som utviklingen av CMOS (Complementary Metal Oxide Silicon) teknologien har gått fremover, har tilgangen på lysfølsomme elementer også blitt bedre, samt dokumentasjon av slike. CMOS teknologi den fordelen at kretser som behandler og beregner enkelt kan implementeres på samme brikke som lysfølsomme elementer i helt standard prosesser. Det er derfor fristende å lage en integrert krets som benytter seg av dette til å gjøre noe billedbehandling direkte på brikken. For ideer til analog billedbehandling går vi til biologien og ser nærmere på synet.

1.3.1 Synet - biologiens optiske krets

Synet består av mye mer enn bare øynene. En stor del (45%) av hjernen går med til synet. Synet er bygd opp som et stort system, der prosessering foregår i flere trinn. Hvert trinn har også tilbakekopling til tidligere trinn. For hvert trinn trekkes mer informasjon ut fra det øyet ser og har sett tidligere. For skarpsynet (i sentrum) kan vi si at øyet ser et mønster, eller rett og slett stillestående objekter. Sidesynet ser hovedsaklig bevegelser.

1.3.2 Mer detaljert om syn

Figur 1 viser et tverrsnitt av retina, som betraktes som en del av hjernen. Den kan deles inn i tre lag: fotoreseptorene (staver, tapper), indre nuklea lag (horisontale celler, bipolare celler og amacrine celler) og overflate laget, som består av ganglion celler (ulike typer). De tre lagene separeres av to lag med synaptiske forbindelser, ytre plexiforme lag (OPL) og indre plexiforme lag (IPL).



Figur 1 - tverrsnitt av retina, forenklet modell, fra WebVision (<http://webvision.med.utah.edu>) med tillatelse. Figuren viser bl.a. staver og tapper (rods, cones), horisontale celler, bipolare celler, amakrine celler og ganglion cellene.

Lyset passerer gjennom hele retina for å nå de lysfølsomme strukturene, stavene og tappene, som er plassert på baksiden av retina. Dette er fordi de trenger konstant forsyning av molekylet retinal, bedre kjent som vitamin A. Når retinal eksponeres for lys endrer det struktur, og elektrokjemisk energi frigjøres. Slik detekteres lys.

Det ytre plexiforme består av synaptiske forbindelser mellom fotoreseptorene og bipolare og horisontale celler. Forskjellige typer celler er tilkoplest på ulike måter. De horisontale cellene er tilkoplest mange reseptorer (mye over lapp, altså at en fotoreseptor er tilkoplest flere celler). De bipolare cellene er ikke tilkoplest like mange reseptorer. I det indre plexiforme laget koples bipolare og amakrine celler til ganglion cellene, som sender ut nervepulser videre inn til synssentrene i hjernen.

Både bipolare celler og ganglion celler deles ofte inn i to hovedgrupper: PÅ og AV, der PÅ cellene er følsomme for lyse objekter mot en mørk bakgrunn og AV cellene for mørke objekter mot en lys bakgrunn. Stavene, som er spesielt brukt ved lavt lys, er kun tilkoplest PÅ celler, som er fornuftig da de skal se når bakgrunnen er mørk (om natten).

Det er store parallelle beregninger som foregår i retina, og man tror at mye filtrering blir gjort før hjernen tar over. Et sentralt begrep er sentrum – omgivelse strukturen. Den består i at noen celler aktiveres ved lys i sentrum (hver celle er tilkopleet mange fotoreseptorer) mens andre aktiveres ved lys i omgivelsene og deaktiveres ved lys i sentrum. Kombinert med PÅ og AV typene som hver av disse kan ha, blir mye informasjon trukket ut fra lyset. Det finnes også celler som er spesialisert til å se spesielle fenomener som f.eks. blinkende lys. Hentet fra *How the Retina Works* (Kolb 2003).

1.3.3 Periferisynet

Som vi ser er synet meget komplisert og sammensatt. Det er lurt å begrense seg til en del av dets funksjonalitet/oppgaver. Jeg ønsker med denne oppgaven å se på muligheten til å etterlikne den bevegelsessensitive delen av synet – nærmere bestemt periferien av netthinnen. Oppgaven er også begrenset til å se på et øye alene.

I studier av menneskehjernen, deler man den gjerne inn i flere lag etter når i utviklingen de forskjellige delene av hjernen oppstod. Som nevnt tidligere, styrer de eldre delene av hjernen grovt sett underbevisste handlinger, og disse er felles for menneskene og mange dyr. De øvre, nyere utviklede deler av hjernen, er spesielle for mennesket og styrer bevisste handlinger og tenkning.

Siden øyet er såpass sammensatt, er det fornuftig å kun se på en del av dets omfangsrike funksjonalitet. Jeg synes det er fornuftig å følge i ”naturens fotspor” ved å begynne med de funksjoner som ble utviklet tidligst; de man regner som underbevisste. For synet er dette funksjoner som navigasjon, reaksjon på bevegelser og reaksjon på lys. Sistnevnte er for eksempel å tilpasse seg til sterkt eller svakt lys. Jeg vil lage et kamera som innehar to av disse funksjonene som man finner i periferien av netthinnen – tilpasning til lysnivået og reaksjon på bevegelser.

1.4 Problemstilling

Inspirert av øynene til mennesker og dyr, som jeg fra nå av refererer til som det biologiske øyet, ønsker jeg å finne svar på følgende spørsmål:

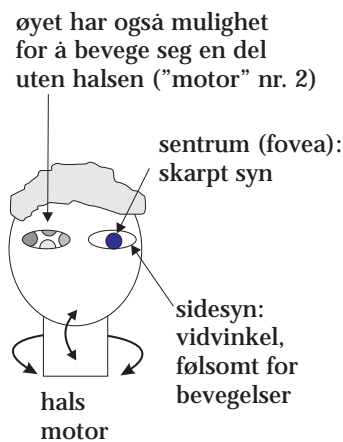
Kan vi (med en standard CMOS prosess) lage et kamera som er f lsoomt for bevegelser slik som periferi synet i det biologiske øyet?

Dette er interessant fordi det er en etterlikning av en av løsningene i biologien. Følsomhet for bevegelser og rapportering av hvor disse er, er en av de mest grunnleggende elementer av syn (eng. vision) ifølge artikkelen ”The human vision system”. Denne typen syn er plassert i de lavere deler av hjernen, altså de områdene av hjernen som ble utviklet tidligst, og som også lavere vesener innehar. Disse områdene av hjernen styrer for det meste underbevisste handlinger, som pust, hjertet, reaksjon på sansestimuli osv. Høyere deler av hjernen styrer bevisste handlinger, handlinger som krever konsentrasjon. Disse delene av hjernen er de som senest ble utviklet, og er spesielle for oss mennesker.

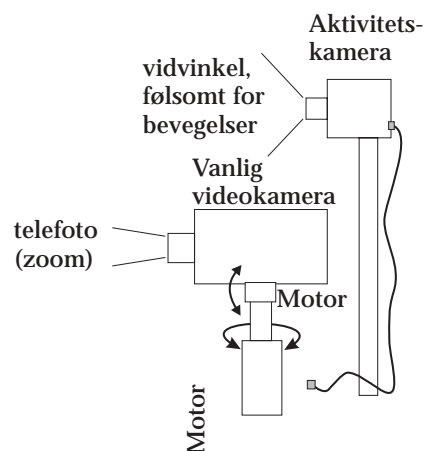
1.5 Et større tenkt system

Kameraet skal, som problemstillingen sier, forsøke å etterlikne visse deler av menneskets sidesyn. Et tenkt oppsett man kan bruke for å demonstrere dette er vist i **figuren under**. Et eller flere kameraer koples opp mot et vanlig videokamera med motor. Kameraet sier i fra hvor noe skjer, i form av x- og y-koordinater. Motoren, med tilhørende elektronikk som forstår nervepulser kodet med AER protokollen, forklart senere, snur så videokameraet i retning der det hender noe. Dette skal demonstrere likheten med oppførselen til mennesker og dyr som merker bevegelser i sidesynet.

Mennesket



Tenkt oppsett



For å gjøre etterlikningen av biologiens virkemåte så komplett som mulig, går kommunikasjonen som nervepulser. Posisjonen som rapporteres av aktivitetskameraet er sentrum av objektet som flytter seg. Dette er sannsynligvis det som også gjøres i biologien, basert på at sentrum av et objekt føles som det mest naturlige stedet å rette skarpsynet mot.

1.6 Praktiske anvendelser

En opplagt utnyttelse av kameraet vil være i overvåkningskameraer. Istedenfor å vise uinteressant informasjon som safen i en bank, vil kameraet kun vise bankranerne. Da kan en vaktmann sitte med øynene halvt lukket eller lese en bok, og likevel registrere når det skjer noe på skjermen som følge av at sidesynet registrerer bevegelse svært godt.

Slike systemer finnes på markedet i dag, fra billige systemer med 60 % korrekte alarmer, til de virkelig dyre systemene med 65 – 70 % korrekte alarmer. Det blir imidlertid en del falske alarmer (30-40 %) særlig utendørs, som langt i fra er godt nok. Disse systemene baserer seg gjerne på metoder der man digitaliserer bildet, og så kontinuerlig skanner inn bildet, som et vanlig videokamera. Prosesseringsenheten, gjerne en tilkople PC som likevel lagrer alle videoopptak, sammenlikner så hver piksel med ny informasjon for å se etter endringer.

2 GENERELT BAKGRUNNSSTOFF

2.1 Strømmen gjennom transistoren

Formelen som er brukt i utregningene av strøm er EKV formelen (Vittoz 1994) vist nedenfor. Denne formelen er valgt pga. god presisjon både i svak og sterk inversjon, selv om den har få parametere slik at den er relativt enkel å bruke. (Det finnes formler for strøm gjennom transistoren som har hundrevis av parametere.) Strømmen gjennom transistoren (*drain* strømmen – I_D) kan uttrykkes ved:

$$I_D = I_F - I_R \quad (1)$$

der I_F er forover komponenten av strømmen og I_R er revers komponenten av strømmen. Disse kan modelleres med akseptabel presisjon for et bredt område av strømmer som (reverskomponenten i parentes):

$$I_{F(R)} = I_S \cdot \ln^2 \left[1 + \exp \left(\frac{V_G - V_{T0} - nV_{S(D)}}{2nU_T} \right) \right] \quad (2)$$

- ▶ V_{T0} er terskelspenningen
- ▶ V_G og V_S er spenningen på *gate* og *source* terminalene
- ▶ n = mellom 1 og 2 er "slope faktor" som brukes for å tilnærme kurven til virkeligheten
- ▶ $I_S = 2n\beta U_T^2$ er spesifikk strøm for transistoren
- ▶ $\beta = \left(\frac{W}{L} \right) \mu C_{OX}$ er transistorens overføringsparameter, der W og L er transistorens bredde og lengde som kan endres av designeren.
- ▶ μ er mobiliteten til elektroner i silisium
- ▶ C_{OX} er kapasitans på tynnoksidet (under gaten) og finnes i prosessbeskrivelsen.

Formelen kan forenkles dersom den skal brukes kun enten i svak inversjon eller sterk inversjon. I utregningene for kretsen er den brukt slik den er presentert i (1) og (2).

Dersom $V_{GS} \ll V_{T0}$ vil transistoren befinne seg i svak inversjon, og (2) kan byttes ut med:

$$I_{F(R)} = I_S \cdot \exp \left(\frac{V_G - V_{T0} - nV_{S(D)}}{nU_T} \right) \quad (3)$$

Formel (3) viser at strømmen for en MOS transistor er eksponentielt bestemt av spenningen.

2.2 Strømmen i biologiske nevroner

Nevronet er en spesialcelle i nervesystemet. Det klassiske synet på nevronenes virkemåte er at de mottar signaler fra andre nevroner. Et nevron er gjerne tilkopleet mer enn tusen andre nevroner via dedikerte forbindelser. Når et nevron blir stimulert, ofte fra andre nevroner, blir kapasitansen over membranet (celleveggen) ladet opp. Hvis nevronet mottar nok stimuli innen en viss tid, vil spenningen nå en terskel. Nevronet sender da ut en nervepuls på axonet, som er en forbindelse til et eller flere andre nevroner. Axonet kan være veldig langt, og er derfor isolert og har såkalte Ranvier noder som repeterer/forsterker opp signalet underveis.

Kjernen til nevronet består av en cellevegg – membranet – med hovedsaklig Na^+ (natrium) på innsiden og K^+ (kalium) på utsiden. Membranet består av molekyler ordnet i en formasjon som kalles tolags-strukturen ("bilayer"). Disse molekylerne er polariserte, akkurat som vannoverflaten, og utgjør dermed en barriere for energi. Svært få ioner slipper gjennom, så vi kan se på membranet som en isolator. Når Na^+ diffunderer ut av cellen gjør dette at det blir en negativ ladning inne i cellen. Dette trekker på ioner fra utsiden, hovedsaklig K^+ , slik at det oppstår en drift inn som følge av det elektriske feltet som oppstår.

Når reverspotensialet, V_R , over membranet tilfredsstiller likningen

$$V_R = -\frac{kT}{q} \ln \frac{N_{in}}{N_{ex}}, \quad (4)$$

vil drift av K^+ ioner inn og diffusjon av Na^+ ioner ut være i balanse. N_{in} er konsentrasjonen av ioner inne i cellen og N_{ex} er konsentrasjon av ioner i væsken rundt.

Netto strøm gjennom membranet (celleveggen) er gitt ved likningen

$$I = (V_K - V)G_K + (V_{Na} - V)G_{Na} + (V_{Cl} - V)G_{Cl} \quad (5)$$

Første ledd beskriver drift av kalium, andre ledd diffusjon av natrium. Det siste leddet er meget lite, og for enkelhets skyld utelatt fra diskusjonen. Konduktansene G_K , G_{Na} og G_{Cl} i (5) er eksponentielt styrt av membranspenningen. Strømmen gjennom celleveggen er derfor eksponentielt avhengig av spenningen.

Når $I=0$ sier vi at nevronet hviler. Det er da polarisert til en negativ spenning, som vi kan se ved å regne ut V_0 – hvilespenningen. Ved å styre spenningen over membranet lages kanaler som kun slipper gjennom en bestemt type ioner, og man styrer dermed membranets permittivitet til Na^+ og K^+ og følgelig strømmen. Dersom potensialet over membranet når en terskel på ca. -40 mV, depolariseres membranet nok til å starte en selvforsterkende reaksjon som genererer en nervepuls. Mens nervepulsen sendes ut er nevronet ikke følsomt for stimuli utenfra.

Det er laget flere kretser som simulerer nevronets oppførsel, fra de helt enkle til store kompliserte kretser som i stor grad oppfører seg likt som nevronet.

2.3 Forskjeller og likheter

De viktigste forskjellene mellom biologien og mikroelektronikk, som vi støter på i denne oppgaven, er at man i biologien har mange flere ledninger. Dette er mulig bl.a. på grunn av en tredimensjonal struktur, og er viktig for biologiens massive beregninger i parallell. Det at like beregninger gjøres flere steder gjør i tillegg til at naturen oppnår stor nøyaktighet, også at biologien er mer robust mot feil.

I elektronikk derimot, er ledningene av en mye høyere kvalitet. Vi kan sende signaler både mye raskere og over større avstander enn det som er mulig i biologien. Når biologien skal etterliknes hadde det likevel vært ønskelig med flere ledninger. Vi må derfor utnytte fordelene vi har i mikroelektronikk til å veie opp for ulempene ved et lite antall ledninger. Dette blir gjort i adresse-hendelse protokollen, beskrevet senere.

Som vi ser, er de grunnleggende prinsippene like for transistoren og for nervesystemet. Det er også verdt å legge merke til det eksponentielle forholdet mellom strøm og spenning for både biologiske nevroner og for transistorer som opererer i svak inversjon.

Dette gjør at det også er en sannsynlighet for at noen av prinsippene i biologien kan overføres til mikroelektronikk.

2.4 Teknologivalg

Det er i dag mange tilgjengelige teknologier for produksjon av integrerte kretser som alle har sine fordeler og ulemper. De mest dominerende er CMOS, BiCMOS, CCD og GaAs. For produksjon av digitale foto- og videokameraer er CCD (Charge Coupled Device) teknologien den vanligste. Til de fleste andre formål er CMOS (Complementary Metal Oxide Silicon) den dominante teknologien, og er både velutviklet og billig.

2.4.1 Fotobrikker i CMOS

Fordelene ved å benytte CMOS teknologi er mange. Teknologien er moden og utviklingen går raskt. Dette er i stor grad fordi CMOS teknologi benyttes i bl.a. prosessorer og minnebrikker som krever hyppig forbedring og nedskalering. Biblioteker med både digitale og analoge kretser er lett tilgjengelig, og det foreligger godt med måledata og konstruksjonsteknikker.

For forskning er det viktig at prototyper kan fremstilles til en fornuftig pris. Noen firmaer har spesialisert seg på å samle opp mange utlegg for integrerte kretser og sende disse samlet til fabrikkene. Dette gjør det er mulig å få en pris som ikke hadde vært mulig ellers. Tilgangen til prototyper gjør også at forskningen ikke kun består av teori. CMOS er forøvrig den billigste prosessen tilgjengelig i dag.

I CMOS prosesser er det dessuten meget enkelt å integrere prosessering og lysfølsomme elementer på samme brikke. Det finnes mange lysfølsomme elementer tilgjengelige i en standard CMOS prosess, som vi skal se på senere.

Det er også et par ulemper ved bruk av CMOS teknologi til fotobrikker. De nyeste prosessene er som regel ikke optimalisert for analog kretskonstruksjon, siden det er digitale prosessorer og minnebrikker som presser utviklingen. Blant annet er forskjellen mellom identiske transistorer relativt stor, men dette kan man delvis se bort fra i digitale kretser. Fotodetektorene er heller ikke standard elementer. Disse er det opp til designeren selv å passe på at fungerer. Visse andre ordens effekter som er nyttige i analog konstruksjon blir også gjerne oversett når utviklingen presses fremover. Effekter slik som at det går an å operere i transistorens ”av” område blir gjerne forsøkt fjernet, fordi de ikke er ønskelige i digitalt design.

2.4.2 CCD og kombinasjonen CMOS/CCD

I motsetning til CMOS prosessen, ble CCD originalt utviklet for analog signalbehandling og videoapparater. Dette har gjort det enklere å lage rene kameraer i disse prosessene, og CCD er den teknologien som i dag er mest brukt og gir best bildekvalitet. For kameraer med en viss intelligens, såkalte smarte sensorer (se side 27), har utvikling gått tregt på grunn av begrensningene i CCD teknologien.

En stor ulempe for analoge kretser i CCD prosesser er klokkingen. Selv enkle operasjoner trenger et stort antall klokkesignaler. I en kombinasjonsprosess CMOS/CCD vil klokkingen av CMOS delen medføre stor støy til de analoge elementene.

CCD prosesser har ikke gode CMOS elementer og CMOS/CCD prosessene har ikke gode CCD elementer. Overføringen av signaler mellom CCD og CMOS elementene på samme brikke krever dessuten spesielle lese- og skrivekretser. Dette tar stor plass på brikken. Store kapasitanter medfører også høyt effektforbruk.

2.4.3 Andre prosesser: BiCMOS og GaAs

BiCMOS prosesser har bipolare transistorer i tillegg til N- og PMOS transistorer. De bipolare transistorene har vært mye brukt i analog design, men på grunn av den store størrelsen og prisen er de lite gunstige til kameraer.

GaAs er kjent for god hastighet for digitale og analoge kretser, men er ikke kommet så langt i utviklingen som CMOS. GaAs er dessuten dyrt og for analoge kretser kan det være et problem at de opererer på veldig lave spenninger. Optisk-elektroniske elementer som brukes i kameraer finnes kun i veldig spesialiserte prosesser.

2.5 Smarte sensorer

Til forskjell fra aktive sensorer gjør smarte sensorer prosessering på et høyere nivå. Det er ikke kun snakk om å gjøre bildet best mulig, men å trekke ut annen informasjon fra bildet. Ved å gjøre dette ”on-chip” mens signalene fremdeles er analoge fremfor i en generell prosessor, kan man i tillegg til fordeler som hastighet og størrelse også trekke ut informasjon som går tapt ved kvantiseringen (omgjøring fra analoge til digitale signaler).

En smart sensor er gjerne bygget opp av følgende tre elementer (Moini 1997):

1. Fotodetektoren, en lysfølsom struktur som gjør om lys til elektriske signaler i form av strøm eller spenning.

2. Fotokretsen, en krets som er bygget rundt fotodetektoren og forsterker opp signalet til et akseptabelt nivå.
3. En smart del som trekker ut annen informasjon enn bare lysintensiteten/fargen. Dette vil være den delen som trekker ut informasjon om kameraet observerer endringer.

2.6 Romlig prosessering

Romlig prosessering er svært krevende, og forenklinger som at man på forhånd vet noe om hvordan bildet kan se ut, er helt nødvendige. Algoritmer som gjør dette blir stadig mer avanserte, men innen feltet visjonsbrikker/smarte sensorer har man heller valgt å følge biologiske modeller. Realisering av beregnende algoritmer i maskinvare er ofte vanskelig, fordi disse algoritmene involverer romlig informasjon fra mange billedpunkter i et område. Siden koplinger generelt er en begrensende faktor, ikke minst fordi det er arealkonsumerende, er beregnende algoritmer generelt regnet for ikke å være særlig egnet for mikroelektronikk. Derimot er biologiske modeller ofte enkle, og benytter informasjon kun fra nabopunktene i matrisen. Biologiske modeller er derfor meget attraktive i implementasjoner i mikroelektronikk.

Smarte sensorer krever ofte prosessering av informasjon mellom nærliggende elementer. En metode som gir forbindelse mellom noder er kjent som *resistive nettverk*. Slike nettverk krever et minimum av koplinger og areal.

Det enkleste resistive nettverket er utjevningsnettverket. Man ønsker da å senke skarpheten i et bilde. Motstandene som er koplet mellom elementene gjør at det kan lekke litt strøm fra lyse noder til mørke noder, og dermed utjevne bildet. Graden av utjevning kan man da kontrollere ved å justere motstandene.

I resistive nettverk har man ofte behov for store motstander. Dette er fordi fotostrømmen er såpass liten, og for å lage en brukbar spenning må motstanden være tilsvarende stor. Passive komponenter i CMOS har ikke stor nok motstand. Det er derfor utviklet en del aktive motstander med langt høyere og regulærbar resistivitet.

I motsetning til CMOS prosesser har ikke CCD prosesser den samme distribusjonsmekanismen for ladninger rundt i kretsen. Derimot må man i CCD prosesser benytte klokking i langt større grad. Dette setter imidlertid en begrensning på hva slags romlig prosessering det er mulig å oppnå, og rene CCD prosesser er ofte begrenset til addisjon og subtraksjon av ladninger. På den andre siden kan utjevning av bildet oppnås ved rett og slett stille på klokkingen.

Det har som generelt har vekket størst interesse er teknikker for adaptasjon og kantforsterkning.

2.7 Temporær prosessering

I tillegg til romlig informasjon vil vi ofte ta inn tiden som en ekstra dimensjon. Den vanligste bruken av temporær prosessering er til bevegelsesdetektering. Verdien av dette kan man se ved å studere insekter, som har et tilsynelatende enkelt synssystem, men likevel er svært presise i sine bevegelser, i stor grad på grunn av bevegelsessensorer i øyet.

Når man arbeider med temporær informasjon er man nødt til å huske tidligere verdier slik at man kan sammenlikne og beregne ut fra dette. Mye av arbeidet med bevegelsessensorer går nettopp på denne biten, altså å huske gammel verdi. Det er to utbredte måter å gjøre dette på. Ved *lagring* og ved *forsinkelse*.

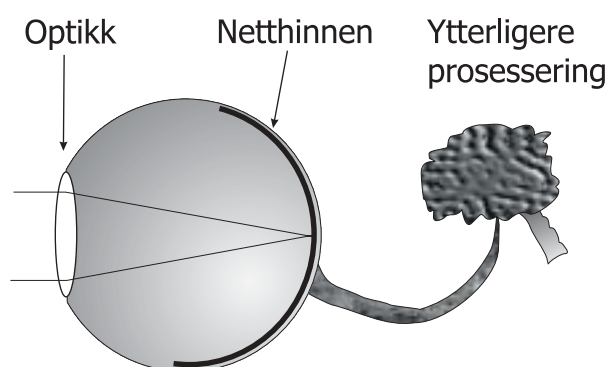
For å se endringer er det nok å innføre en forsinkelse, som i analog design krever en stor kapasitiv node til å holde på informasjonen. I motsetning til lagring av signalet injiseres og utleses det kontinuerlig og sammenliknes med et signal uten forsinkelse.

Det er særlig bevegelsesdetektering man arbeider med i dette feltet, men det er også laget kretser for fokusering, måling av endring i bildet, fart, beregning av tid til kollisjon, retning og avstandsmåling.

3 HELE SYSTEMET

3.1 Inndeling av systemet i tre deler

En forenkling av synet, delt inn i tre deler er i Figur 2. Lyset kommer først i kontakt med optikken, som fanger opp lyset og forminsker bildet. Lyset treffer så netthinnen/retina, som gjør om lyset til elektrokjemiske signaler. De forskjellige områdene av netthinnen har ulike egenskaper. Periferien er spesielt følsom for endringer. De elektrokjemiske signalene sendes videre til ytterligere prosessering i de forskjellige synssentrene i hjernen. Sidesynet og skarpsynet går helt separate veier videre (sidesynet fra de to øynene samles og sendes videre til et sted, og skarpsynet samles og sendes til et annet sted. Det er likevel koplinger mellom disse).



Figur 2 - en forenklet modell av synet vårt, delt inn i tre deler

3.1.1 Optikk i øyet og kameraet

Den delen av øyet som først mottar lyset er linsen (rettere sagt hornhinnen, som også fungerer som en linse), og dermed et naturlig sted å starte. Den har som oppgave å forminske bildet ned til størrelsen av netthinnen. Øyet har en konveks linse, som er en linse som er tykkest på midten. En slik linse samler lyset, og brukes til å forminske bildet. Hvor mye en gitt linse forminsker et bilde, er avhengig av kurven på linsen. En tykk linse forminsker mer, og dette gjør at netthinnen/sensoren ser mer av virkeligheten. Et mål på hvor mye en linse forminsker er vinkelen, der store vinkler forminsker mer. Synet i periferien av netthinnen utnytter øyets store vinkel for å være effektivt. Skarpsynet derimot, er konsentrert i midten, og bruker kun en vinkel på noen få grader.

Kameraer har også vanligvis en linse, selv om det er fullt mulig å lage et kamera uten. Man lar da alt lys inn til sensoren gå gjennom et svært lite hull, og får da et fokusert bilde. Det knøttlille hullet vil naturligvis gjøre at svært lite lys slipper til, slik at kameraet kun fungerer i sollys. Uten linse vil man heller ikke få noen forminskning av bildet, og store objekter som mennesker og biler vil bare kunne observeres på lang

avstand. Siden jeg etterlikner synet i periferien anser jeg det for nødvendig med en linse i kameraet og at denne har vid vinkel.

3.1.2 Linseholder (kamerahus)

Rent praktisk trengs en linseholder for å holde linsen på plass over sensoren. Linseholderen er ikke særlig biologisk inspirert, og har få likheter med øyet. Det kan nevnes at den er sort innvendig for å begrense refleksjoner. Netthinnen i øyet inneholder stoffet melanin som gjør den sort i glippene der det ikke finnes staver eller tapper. Her slutter likhetene med øyet, som bl.a. er fylt med en geleaktig væske fremfor luft, og som har muskler som kan vri øyet i x- og y-aksen og rotere det.

De viktigste egenskapene linseholderen må inneha er at den er lystett, slik at man unngår strølys, og at avstanden fra bakstykket på objektivet til sensoren er korrekt. Dette for at brennpunktet skal treffe akkurat på sensoren slik at bildet er fokusert. Ellers må den selvfølgelig passe til objektivet og kunne festes rundt den integrerte kretsen på en grei måte.

Det er fint om posisjonen til linsen kan varieres, slik at de lysfølsomme elementene ikke trenger å være plassert midt på brikken. Muligheter for justering gjør det mer sannsynlig at linseholderen kan brukes om igjen i fremtidige prosjekter. Justering av linsens avstand fra sensoren gjøres med fokus på objektivet. Andre justeringer må bygges inn i linseholderen. Et siste krav til linseholderen er at den må være enkel å bygge.

3.1.3 Netthinnen og sensoren

Den andre delen av systemet i Figur 2 er netthinnen, i vårt tilfelle periferien av den, og det er her hovedtyngden av oppgaven ligger. Netthinnen gjør om lys til elektrokjemiske signaler, og cellene i periferien er særlig følsomme for endringer. Den tilsvarende delen i kameraet kalles for sensoren, og på grunn av dens evne til å føle endringer faller den inn under begrepet en smart sensor, beskrevet tidligere (avsnitt 2.5).

3.1.4 Ytterligere prosessering

Den tredje delen av systemet har jeg kalt ytterligere prosessering, og gjøres steg for steg i de ulike synssentrene i hjernen. Hvert nivå med prosessering tar oss lenger vekk fra det observerte lyset, og forteller oss mer om hva vi faktisk ser. Tidlige nivåer detekterer kontraster, endringer og slikt. Videre analyseres bl.a. informasjon om kanter og bevegelse. I enda høyere nivåer registreres det hva vi ser, for eksempel ansikter.

Et kjennetegn ved hvordan nervesystemet er bygd opp, er at et nevron får impulser fra flere tilkoblede nevroner, og kombinerer informasjon fra disse. Måten informasjonen kombineres på kan variere, men er som regel basert på at nevronet må få en viss mengde stimuli innenfor et gitt tidsrom. Når en viss mengde stimuli mottas, gir nevronet beskjed ved å sende ut en nervepuls. For en gitt gruppe nevroner kan det være av interesse i hvilken rekkefølge nevronene aktiveres.

Den ytterligere prosesseringen som skal gjøres i kameraet, vil være å finne ut hvilken posisjon et objekt, som beveger seg, har i bildet. For å gjøre dette må informasjon fra

de smarte sensorene kombineres slik at man kan skille ut hva som er et objekt og hva som er små ubetydelige endringer. Det vil også være naturlig å for eksempel beregne sentrum/størrelse av objektet man ser.

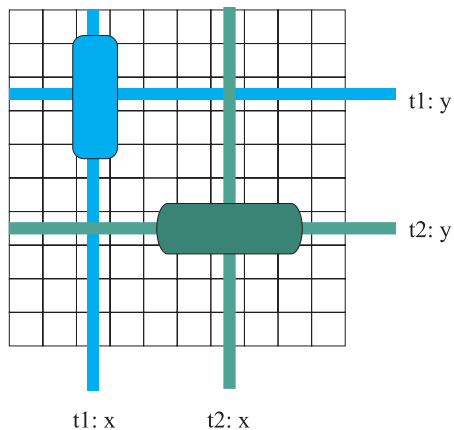
3.2 Utforming av hele systemet

Beslutninger som gjelder generelt for hele systemet er diskutert her.

3.2.1 Et eller flere objekter

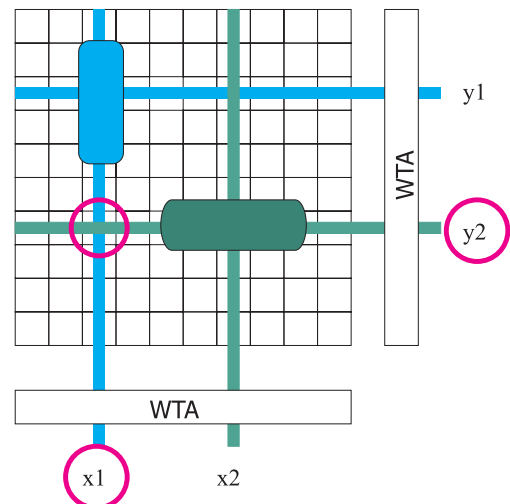
Et viktig valg var å bestemme om kretsen skulle kunne bestemme posisjonen til et eller flere objekter. Et kamera som kunne følge flere objekter viste seg for å bli alt for komplisert.

Med WTA nettverk i 2D blir det riktig



I dette tilfellet har man et todimensjonalt WTA nettverk over hele matrisen. Midterste nevron i objekt 1 sender ut en puls ved tiden $t1$. Tilsvarende for objekt 2 ved tiden $t2$. Det er lett å skille i to koordinat-par, et for tiden $t1$ og et for tiden $t2$.

WTA langs kantene klarer kun et objekt



Det er viktig at x- og y-koordinatene hører sammen. Dersom man har et vanlig WTA for hver akse, vil det kunne forekomme at to objekter, som vist, vil kunne vinne på hver sin akse, og koordinatparet vil være satt sammen av x-koordinaten til det ene objektet og y-koordinaten til det andre objektet.

Mulige kombinasjoner:

$(x1, y1)$, $(x1, y2)$, $(x2, y1)$, $(x2, y2)$

Figur 3 - En krets som skal kunne følge flere objekter, trenger en todimensjonal winner-take-all krets (WTA), som vist til venstre. Med to WTA kretser langs kantene, vist til høyre, kan flere objekter føre til feil.

En krets som kan følge flere objekter blir svært mye mer komplisert enn kameraet i denne oppgaven. Mye er et resultat av at et ekte todimensjonalt winner-take-all nettverk behøves, som illustrert i Figur 3. De viktigste forskjellene er:

- ▶ En krets som følger flere objekter må ha en todimensjonal winner-take-all krets med et nevron pr. piksel istedenfor et nevron pr. rad og kolonne. Når et objekt i bevegelse er funnet, vil det midterste av de nevronene som registrerer objektet, sende ut en puls på raden og kolonnen den er tilkople. Dette er skissert i Figur 3.
- ▶ En annen utfordring, som flere-objekts kretsen må ta hånd om, er at den må kunne bestemme hvilke piksler som observerer samme objekt. Dette er ikke helt trivielt, da et objekt i bevegelse sjelden vil ha endringer i hele flaten (alle piksler) og lett kan oppfattes som flere objekter. En strategi for å finne ut hvilke piksler som tilhører et objekt, er å forsøke å finne omrisset ved å se etter kanter. Dette vil sannsynligvis virke godt på objekter med store uniforme flater, og mindre godt for objekter med mange indre kanter (for eksempel en stripeball). I simuleringer er flere metoder forsøkt for å finne kanter med stor sikkerhet, bl.a. såkalte "fuse" kretser.
- ▶ Beregning av sentrum blir vanskeligere for flere-objekts kretsen, siden beregningen gjøres i to dimensjoner. I kretsen som begrenser seg til å følge et objekt, gjøres en "pseudo" todimensjonal beregning, ved at man finner raden og kolonnen med mest endring og så kopler dette sammen til en (x, y) koordinat.
- ▶ Winner-take-all (WTA) kretsen må kunne tillate flere vinnere, dersom flere objekter skal kunne detekteres. Dette gjøres vanligvis ved å la vinneren kun holde nevroner i en viss avstand fra seg selv nede, slik at flere nevroner kan vinne, dersom de har en viss innbyrdes avstand. Denne avstanden må ikke være for kort, da man kan få to vinnere som representerer samme objekt, og heller ikke for lang, da et objekt forhindrer andre objekter fra å oppdages.
- ▶ I flere-objekts følgeren jeg arbeidet med (på simuleringer), var denne avstanden variabel – styrt av objektenes størrelse. Mens vinnerens "revir" vanligvis styres av motstander, ble utstrekningen i dette tilfellet styrt av kant-detektorer..

Som listen over viser, blir et kamera som kan følge flere objekter for komplisert for en hovedoppgave. Kameraet er derfor begrenset til å kun følge et objekt. En krets som følger flere objekter kan være en tenkt utvidelse.

3.2.2 Ekte bilde eller bare "meta" informasjon

Ulike steder på netthinnen har ulik sammensetning av celler. Som kjent har mennesker bedre skarphet (oppløsning) i sentrum (eng. fovea) av netthinnen enn langs kantene. Det er denne delen av synsfeltet vi benytter til kompliserte oppgaver (for synet) som for eksempel å lese eller å studere karakteristikken til objekter vi møter i naturen.

Siden kameraet skal etterlikne periferisynet, er det ikke veldig interessant å vise et ekte bilde, for selv om det også i periferien finnes celler som er følsomme for farger, så er det bevegelsesfølsomheten som er mest fremtredende. Å la kameraet kun sende ut

informasjon om endringer er dessuten en god filtrering, og gjør det enklere å tilkople (hypotetisk) utstyr.

Spørsmålet er da hva slags informasjon kameraet skal sende ut. Sentrum av objektene er en god kandidat, siden det er forholdsvis enkelt å beregne, videre bruk av informasjonen vil være enkel, og det kan være sannsynlig at det samme gjøres i øyet. Vurderte alternativer er å overføre omrisset, eller kombinasjonen av sentrum pluss størrelsen. En mulighet var også å se etter retningen av en bevegelse.

3.2.3 Beregning av sentrum

Beregning av sentrum for et enkelt objekt kan gjøres relativt enkelt dersom resultatet ikke behøver å være helt korrekt. Sentrumsberegningen kan gjøres fra kantene av matrisen, som illustrert i Figur 3. Sentrum i begge retninger (rader og kolonner) beregnes individuelt og settes sammen.

Forsøk på å beregne et nøyaktig gravitasjonssentrum medfører mer arbeid enn et raskt estimat, og det er ikke garantert at beregningen blir så nøyaktig fordi et objekt i bevegelse ikke alltid vil generere endringer over hele flaten (f.eks. et stort ensfarget objekt). Det kan derfor være mer fornuftig med en omtrentlig sentrumsberegning.

Den omtrentlige beregningen gjøres ved at man ser etter raden/kolonnen med mest endringer fremfor å beregne sentrum på grunnlag av hele objektets form. Dette gjøres ved å benytte to winner-take-all kretser, en for radene og en for kolonnene. Disse sammenlikner summen av endringer for hver rad/kolonne og plukker ut raden/kolonnen med størst sum av endringer (hver piksel med endringer vil bidra til en større total sum for raden og kolonnen den er tilkople).

Avhengig av objektets form, kan dette bli en svært god eller svært dårlig beregning. For objekter som er tykkest på midten, og genererer endringer i hele dette området, vil også kameraet plukke ut midten som sentrum. Et eksempel er en ball (ikke ensfarget). For objekter som er tykkest på en side, som f.eks. en trekant, vil den ene siden bli det beregnede sentrum, som jo er feil. I begge tilfeller vil sentrum uansett være et sted på objektet, som ikke er så aller verst.

3.2.4 Utlesningsmekanisme

Når sensorer (som billedpunkter) skal avleses, er det to grunnleggende prinsipper man følger: Skanning og hendelsesbasert avlesning.

Skanning vil si at man jevnlig avleser verdien til alle sensorene etter tur, og er den vanlige formen for avlesning av bildesensorene i digitale foto- og videoapparater. Det er vanlig å la kondensatorer i sensorene lade seg opp en viss tid, for deretter avlese verdien og nullstille kondensatoren. Sterkt lys gjør at sensoren blir ladet opp raskere enn svakt lys, og man kan derfor se på spenningen hvor sterkt lyset har vært i gjennomsnitt siden siste avlesning. I videokameraer vil man vanligvis avlese hvert punkt 25-30 ganger pr. sekund. Billedpunkter i CCD teknologien er spesielt egnet til å avleses ved skanning.

Skanning er enkelt å implementere på grunn av sin enkle sekvensielle struktur, og er et klart førstevalg dersom man ønsker å avlese alle sensorene jevnlig. Et problem for følsomme analoge sensorer er at skanning krever klokking, som kan forstyrre sensorene i og med at det forårsaker en del støy.

Hendelsesbasert avlesning er fundamentalt forskjellig ved at det ikke finnes noen ytre mekanisme som styrer avlesningen. Det er billedpunktene selv som bestemmer når de ønsker å bli avlest, og de sender da vanligvis ut en forespørsel om avlesning, men det er også tilfeller, som kameraet, der billedpunktene sender ut sin verdi direkte. Det er ikke nødvendig med klokke, men ofte med en strategi for å håndtere kollisjoner som kan oppstå.

Hendelsesbasert avlesning har sin store styrke i forhold til skanning når det ikke er nødvendig å avlese alle billedpunktene hele tiden. Eksempler kan være når statisk informasjon er uinteressant. Skanning vil da lese ut mye uinteressant informasjon som igjen vil være bortkastet båndbredde.

En forskjell som er mindre opplagt er at hendelsesorientert avlesning ikke kan bruke "sample-and-hold" fotokretser, fordi disse kretsene må avleses jevnlig. Hendelsesorientert avlesning krever at fotokretsene må kunne avleses når som helst, der variabelen som angir lyset varierer kontinuerlig, fremfor sensorer som avleses til bestemte tider styrt av en klokke.

For kretser som ser etter endringer kan dette være spesielt nyttig. Endringen vil oppdages umiddelbart etter at den inntreffer, og et signal kan sendes ut. Med skanning må man derimot vente på at billedpunktet skal bli avlest, og verdien vil dessuten være gjennomsnittet av lyset for en viss tid. Bevegelsessensorer som skannes er derfor helt avhengige av at oppfriskningsfrekvensen, hvor ofte det skannes, er høy nok. En utfordring ved høye oppfriskningsfrekvenser er at sensorene får mindre lys på seg før de blir nullstilt. Dette forverrer signal til støy forholdet. I tillegg vil klokken gå raskere.

For bevegelsessensitive kameraer er konklusjonen at det er mange fordeler ved hendelsesorientert design. Spart energi og båndbredde ved at mindre informasjon overføres. Økt hastighet får vi ved at et billedpunkt straks kan registrere og sende ut sin verdi. Til sist får man økt følsomhet ved at reelle verdier brukes og ikke gjennomsnitt. Dette er også det som gjøres i biologien.

3.2.5 Nevromorf koding av utgangssignalene

En opplagt grunn til å benytte nevro-morf koding (se avsnitt 6.1 side 49) i kameraet er å etterlikne biologiens egen form for kommunikasjon, hvilket gjør kretsen mer biologisk inspirert. Nevromorfe signaler har noen egenskaper som er gode for kameraet:

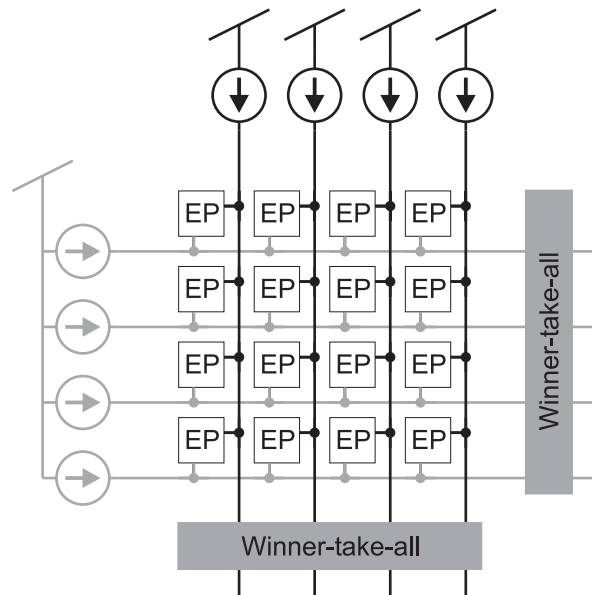
- ▶ Analog i tid: Siden kameraet er hendelsesorientert, er dette en svært fordelaktig egenskap. Kameraet kan da når som helst sende ut sin output, uavhengig av digitale klokkesignaler.
- ▶ Digital i verdi: Vi er kun interessert i hvor noe beveger seg, en eksakt angivelse av rad og kolonne. En digital verdi (adressen til nevronet) er dertil svært godt egnet.

Jeg har valgt å la utgangssignalene bruke adresse-hendelse protokollen (se avsnitt 7.2 side 67), som gjør det enklere å kople sammen kretsen med andre kretser som kan lese slike signaler. Det er en fordel at det finnes utstyr på laben som kan kommunisere via denne protokollen, og hente signalene inn i en datamaskin der man kan analysere data fra kretsen. En siste motivasjon for å bruke nervepulser som

utgangssignaler er å teste ut en ny winner-take-all krets som er nærmere beskrevet i avsnitt 6.3 på side 51.

3.3 Oversikt over hovedkretsen

Kameraet skal se objekter som beveger seg, følge dem, og rapportere hvor de befinner seg til en hver tid. Utgangen av kretsen er adresser til rader og kolonner etter hvert som objektet flytter seg til en ny rad eller kolonne. Kombinasjonen av en rad og en kolonne bestemmer entydig en piksel som skal vise sentrum av objektet. Adressene som sendes ut av brikken er binært kodet, som spesifisert av adresse hendelse protokollen. Kameraet er begrenset til å følge kun et objekt, da flere objekter blir for komplisert. En oversikt over kretsen, forminsket til 4 x 4 piksler er vist i Figur 4 nedenfor.



Figur 4 - oversikt over kretsen for 4 x 4 piksler.

Som Figur 4 viser, består kameraet av en matrise av endringsdetektorer, merket "EP" (endringspiksel). Figuren viser 4x4 piksler, men kameraet har 42x51 piksler, som totalt utgjør 2142 lysfølsomme punkter. Hver av disse er tilkopleet en rad og en kolonne, som hver har en strømkilde på toppen. Radene og kolonnene er tilkopleet hver sin winner-take-all (WTA) krets. De to WTA kretsene er igjen tilkopleet kretser som håndterer adresse hendelse protokollen, men dette er ikke vist i figuren.

Kameraet fungerer slik:

1. Hver endringspiksel (EP i figuren) ser sin del av det fokuserte forminskede bildet som kommer fra linsen. Endringspikslene sammenlikner lyset den ser med lyset for en stund siden, avgjør om det har skjedd en endring i sin del av det totale bildet, og slutter i så fall å trekke strøm.

2. Et objekt av en viss størrelse vil sees av flere endringspiksler, som da ser hver sin del av objektet. Hvis objektet beveger seg, vil da en gruppe EP-er detektere endring og slutte å trekke strøm. Dette vil føre til at det blir en overskuddsstrøm fra strømkildene som leverer strøm til de radene og kolonnene gruppen av EP-er tilhører.
3. Winner-take-all kretsene vil så velge ut hvilken rad og kolonne som har størst overskuddsstrøm, som vil være raden og kolonnen flest av EP-ene i gruppen tilhører. I de fleste tilfeller vil dette bestemme et omtrentlig sentrum av objektet.
4. Den valgte raden/ kolonnen sendes ut av brikken binært kodet ifølge adressehendelse protokollen, forkortet AER (avsnitt 7.2).

Kretsene som kameraet består av er forklart i egne kapitler. Disse er:

- ▶ Lysfølsomt element og fotokrets, beskrevet i kapittel 4.
- ▶ Korrelasjonskrets og endringspiksel, beskrevet i kapittel 5.
- ▶ Nevroner og winner-take-all kretsene, beskrevet i kapittel 6.
- ▶ AER kretsene, beskrevet i kapittel 7 sammen med resultater fra kameraet i sin helhet.

3.4 Dedikerte testkretser

I tillegg til hovedkretsen er det lagt inn 6 dedikerte testkretser på brikken med mange målepunkter. Disse er:

- ▶ korrelasjonskrets
- ▶ endringspiksel uten fotodiode (erstattet med transistor)
- ▶ ”integrate-and-fire” nevron
- ▶ fire ”integrate-and-fire” nevroner i et ”nevro-WTA”
- ▶ fotokrets m. fotodiode
- ▶ endringspiksel med fotodiode

Av disse kan de fire øverste testes uten lys, altså med lokk på kretsen. Siden disse er enklest å teste, ble disse prøvd først. Målingene er presentert sammen med beskrivelsen av sine respektive kretser.

4 LYSFØLSOMT ELEMENT OG FOTOKRETSEN

4.1 Lysfølsomme elementer/strukturer i CMOS

En fotosensor består av mange lysfølsomme elementer som jeg har valgt å kalle fotodetektorer. Hver av fotodetektorene er gjerne tilkoplek kretser som forsterker og forbedrer signalet. Disse kretsene kaller jeg fotokretser. Mange fotodetektorer med tilhørende fotokretser settes sammen til en matrise der hver fotodetektor ser en del av bildet, og refereres av den grunn ofte til som et billedpunkt eller en piksel.

4.1.1 Fotodetektorer

Fotodetektoren er den viktigste delen av fotochipen, iallfall når man tenker på kvaliteten av innlest data. Dette er det elementet som er i direkte eksponert for lyset, og hvis ikke denne genererer gode data er det vanskelig for senere prosessering å gjøre en god jobb. Det er derfor viktig å ta seg god tid med utformingen av fotodetektorene. En annen vinkling er selvfølgelig å benytte eksisterende fotodetektorer som er kjent for å ha en god virkningsgrad.

Hvis man benytter en standard prosess (uten spesielle fotodetektorer) er det parasittiske effekter vi utnytter. Fotodetektorene i CCD prosesser er de mest velutviklede fordi etterspørselen etter CCD kameraer har vært stor, og etterspørselen etter CMOS og GaAs kameraer har vært tilsvarende liten. Potensialet for CMOS er likevel stort. Her er noen fotodetektorer som kan lages i standard prosesser:

- ▶ Brønn til substrat diode
- ▶ Diffusjon til brønn diode
- ▶ Kombinasjon av de to over, altså en diffusjon til brønn diode og en brønn til substrat diode i parallell for å utnytte arealet bedre.

Fotodioder koples i sperreretning, slik at det kun går en reversstrøm gjennom dem. Når de eksponeres for lys, vil størrelsen av reversstrømmen modelleres av lyset fordi fotoner absorberes og slår løs elektroner i deplesjonssonen, slik at det blir flere ladningsbærere. Den modellerte reversstrømmen kalles fotostrømmen. Effektiviteten til en fotodetektor måler man ved å se hvor godt fotodetektoren absorberer lys av ulike bølgelengder. Tidligere har kortbølget lys vært et problem i CMOS prosesser, men ettersom teknologien har krympet går også dette an å måle rimelig effektivt.

I en vertikal diode består fotostrømmen av to elementer. (1) Drift av elektroner og hull i utarmingsområdet og (2) diffusjonsstrømmen av ladningsbærere utenfor utarmingsområdet. Se Figur 5 b) der strømmen går i reversretning gjennom den viste dioden.

En horisontal diode får fotostrøm ved at lyset frigjør ladningsbærere i brønnen mellom kontaktpunktene, slik at det går en strøm der. I tillegg utgjør substratstrømmen en viktig del av fotostrømmen for horisontale fotodioder.

- ▶ Vertikal bipolar transistor

► Horisontal (toveis) bipolar transistor

I bipolare transistorer vil lyset modellere base strømmen, ved å frigjøre ladningsbærere på samme måte som for fotodioder. Kollektorstrømmen, som vi måler, vil være forsterket med en faktor β som er transistorens forsterkning, og som typisk kan ha en verdi over 100. Dessverre vil også støy forsterkes på lik linje med signalet, så signal til støy forholdet er uendret. Vertikale bipolare transistorer i CMOS prosesser vil alltid ha sin emitter terminal tilkopleet jord fordi *emitteren* er substratet.

► Fotogate

Fotogaten har hentet inspirasjon fra CCD prosesser. Det er ganske enkelt en MOS kondensator som utsettes for lys. Frigjøringen av ladningsbærere gir en spenning over kondensatoren som senere kan avleses. Den har derfor til forskjell fra diodene en integrerende effekt, og må avleses med jevne mellomrom (det går ingen kontinuerlig strøm ut av fotogaten). Til fotogaten hører en transistor som brukes til å åpne og lukke for avlesning og nullstilling av spenning over kondensatoren. Et problem med fotogaten er at polysilisiumet blokkerer en del lys. Løsninger på dette har vært å lage hull i gaten, så lys kan slippe gjennom til substratet/den fotogenererende regionen.

4.1.2 Utfordringer for fotodetektorer

Skalering av CMOS prosessene til stadig mindre mål fører til at absorpsjonsevnen går mot kortere bølgelengder. Langbølget lys går tvers gjennom. Selv om det går an å lage større fotodetektorer horisontalt er det lite man kan gjøre vertikalt. Simuleringer viser at på grunn av dette er det diffusjon-substrat og brønn-substrat dioder som dekker synlig lys best.

I moderne prosesser er både gate- og diffusjonsområder ”silicided”. Silicide blokkerer det meste av synlig lys. Man må derfor passe på å maskere ut silicide i utlegget for de områdene man ønsker skal være lysfølsomme, eller i det minste lage hull, slik at lys slipper til.

Et annet problem i fotobrikker er forskjellen mellom transistorer som er ment å være identiske. Når vi ser på fotodetektorene er dette imidlertid ikke noe stort problem. For det første er fotodetektorer gjerne store, og dessuten tar de ofte strømmen sin fra substratet, som generelt er mer jevnt over det hele enn overflate materialene. For eksempel er standardavviket kun 2% i bipolare transistorer med størrelse $100\mu^2$. Fotodetektorer som benytter brønn fremfor substrat får et høyere avvik.

4.2 Utforming av lysfølsomt element

Som beskrevet er det flere strukturer/konstruksjoner i standard CMOS prosesser som kan benyttes som lysfølsomme elementer. Selv om andre strukturer har mer strøm, er fotodiode valgt, på grunn av dens enkelhet og mindre følsomhet for bl.a. digital støy enn andre elementer. Fotodioden er også den mest vanlige, og standard

lysfølsomt element i den valgte fotokretsen. Den trenger ikke nullstilles, som er et av prinsippene vi følger når vi konstruerer kretser som etterlikner retina.(Bohaen 1996)

Det er forskjellige typer fotodioder, og disse har forskjellig følsomhet for ulike bølgelengder av lyset, men dette har lite å si for kameraet. Veileder hadde tidligere laget en brønn til substrat fotodiode, som gjorde at valget falt på denne.

Størrelsen av dioden, og hvilke strømmer som skulle simuleres med ble resonert på følgende måte. I en fotodiode er det reversstrømmen som modelleres av lyset, så denne kan man bruke som et utgangspunkt for strømmen i totalt mørke. Simulatoren har en modell for substratdioder i den aktuelle prosessen, som ble brukt som modell for fotodioden.

Lyset kan variere med 10^{12} ganger. Veileder mente at en variasjon på 10^6 ganger ville være sannsynlig for en fotodiode, og at en variasjon på 10x for et gitt lysnivå ville være et godt estimat.

Simuleringene ble utført der fotodioden modelleres av en diode og en strømkilde i parallell, der diodens reversstrøm er brukt for totalt mørke, og reversstrømmen $\times 10^6$ for sollys. Simuleringene ble utført for flere nivåer, der hvert nivå hadde en variasjon i strømmen med 10x for å simulere objekter som beveger seg forbi fotodioden.

Siden fotostrømmen er såpass liten (en modellert reversstrøm), er den også veldig følsom for støy. Det er essensielt å holde fotostrømmen godt isolert fra andre signaler, særlig digitale signaler som lett støyer for andre kretser. Siden fotostrømmen må forsterkes i stor grad, er det viktig at forholdet mellom signal og støy er så godt som mulig.

For å oppnå dette er fotodioden laget ekstra stor, og digitale kretser (nevronenes pulser er digitale) er plassert langt vekk fra fotodiodene. Analoge og digitale kretser har også separat forsyningsspenning og jord. Fotodiodens store størrelse gjør at fotostrømmen blir større i forhold til naturlig (termisk) støy og ikke trenger å forsterkes like mye som en mindre strøm. Større lysfølsom overflate gjør at fotodioden tar opp mer lys, noe som særlig minsker støy ved lite lys. I tillegg blir det bedre samsvar mellom fotodiodene på kretsen, da geometriske ulikheter som følge av prosessfeil i produksjonen blir mindre fremtredende, fordi prosessfeilen er uavhengig av størrelsen på dioden og derfor blir relativt mindre når størrelsen på dioden øker. En stor fotodiode gjør at kameraet trenger mindre lys for å se, men på den andre siden får det lavere oppløsning. Også i biologien er god lysfølsomhet prioritert fremfor høy oppløsning i periferisynet.

En fordel med store fotodioder som ofte overses er den optiske filtreringen lav oppløsning gir. Et eksempel på dette er hvis man har et tre i bildet, og hver piksel ser flere blader. Hvis vinden blåser og bladene beveger seg, vil noen reflektere mer lys og noen mindre, slik at hver piksel i gjennomsnitt ikke ser store endringer/forstyrrelser i lyset.

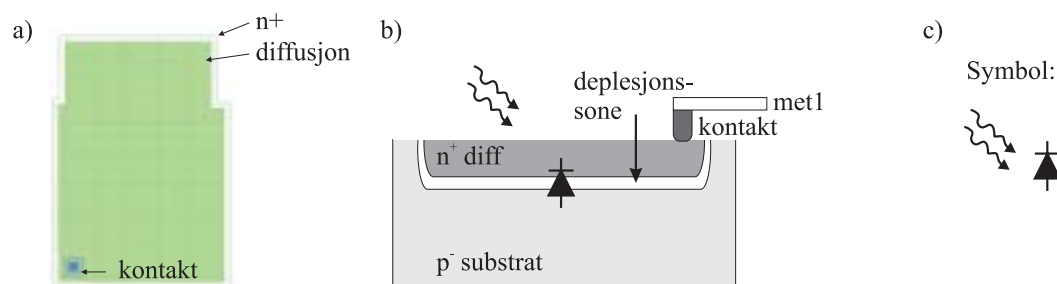
Den konkrete størrelsen av fotodioden ble i utgangspunktet bestemt av en fotodiode veileder tidligere hadde laget. Denne var riktignok testet med en LED (lys

emitterende diode) og ikke vanlig rombelysning. Fotodioden ble derfor laget noe større enn denne, for å være sikker på at lysfølsomheten ble god nok.

4.3 Utlegg av fotodioden

Fotodioden er vist i Figur 5 a. Den lages med to masker, n^+ doping og diffusjon, som til sammen blir en n^+ diffusjon. Den har en kontakt til metall 1 som er tilkoplingen til katoden. Anoden er tilkoplest substratet. Figur 5 b viser fysisk tverrsnitt av fotodioden. Overgangen mellom diffusjonen og substratet utgjør en np diode, altså en diode i sperreretning. Det er en depleksjonssone mellom lagene. Frie ladningsbærere i dette området økes når dioden eksponeres for lys, og følgelig øker også reversstrømmen fra diffusjon til substrat. Figur 5 c viser et vanlig symbol for en fotodiode.

Størrelsen av fotodioden (n^+ diffusjonen) er $11,6 \mu\text{m} \times 17,3 \mu\text{m}$ på det bredeste. Arealet er $\approx 194 \mu\text{m}^2$ minus arelaet dekket av kontakten. Det er meget viktig at fotodioden ikke tildekkes i utlegget, fordi den må eksponeres for lys. Dette gjør ruting av ledninger litt vanskeligere, da de må rutes rundt. På den andre hånd må alle andre kretser tildekkes, fordi eksponering av lys også vil påvirke disse.

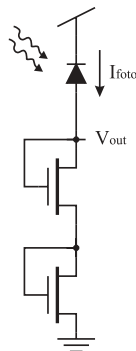


Figur 5 - utlegg av fotodioden, fysisk tverrsnitt som viser depleksjonssonen som lyset modellerer, og krets symbolet

4.4 Fotokretser

Fotokretsen er den kretsen som først mottar fotostrømmen. Den er derfor meget viktig. Fotokretsens oppgave er som regel å forsterke fotostrømmen, som er meget liten, i tillegg til å kompensere for fotodetektorenes lave dynamiske spennvidde. Lyset utenfra varierer med over 10 dekadere, og dette må kompenseres for. Ved at fotostrømmen forsterkes og de store variasjonene reduseres på et tidlig stadium kan øvrige kretser lages mindre følsomme og ha mindre dynamisk spennvidde.

En enkel måte å konvertere fotostrømmen til en spenning på er å bruke en eller flere enkle MOS dioder. Spenningen over en diode er proporsjonal med logaritmen av strømmen, som gjør at man får en stor kompresjon av fotostrømmens store variasjon. Et eksempel med to MOS dioder (diodekoblede transistorer) er vist i Figur 6.



Figur 6 - en enkel fotokrets som benytter diodekoblede transistorer (her to stk) til å logaritmisk komprimere signalet.

En slik fotokrets er svært liten og har stor dynamisk spennvidde, og er derfor mye brukt. Ved lavt lysnivå er den imidlertid treg. Sammentrykningen av input signalet kan være et problem hvis man ønsker å skille signaler fra hverandre. Særlig bevegelsessensorer som leter etter kontraster kan få problemer med slike fotokretser.

For å få bedre dynamisk respons er det essensielt å kompensere for den kapasive lasten inn på fotodetektoren. Det er flere muligheter for å gjøre dette. Et nøkkelord er tilbakeføring. Forskjellige løsninger som baserer seg på buffere og forsterkere gjør en god jobb med å forsterke strømmen, men de har et logaritmisk spennings-til-strøm-forhold, som gjør at kontraster svekkes. En gruppe fotokretser kalles adaptive, som vil si at de tilpasser seg lysnivået. Endringer innen det innstilte lysnivået vil gi store utslag. Dersom lysnivået endres, vil slike kretser tilpasse seg det nye lysnivået.

4.4.1 Sample-and-hold fotokretser

De aller fleste fotokretser er basert på integrasjon. Det vil si at man tar lysintensiteten over et tidsrom og beregner gjennomsnittet. Ved å bruke en metode kalt "sample-and-hold" tar man en kopi av den oppsamlede spenningen, og nullstiller så spenningen i fotodetektorens kondensator. Fordelene med denne teknikken er lineære overføringskarakteristikker, enkelt kontrollerbar lysfølsomhet (ved å endre integrasjonstiden) og liten følsomhet for forskjell mellom transistorer. Ulempen er at man ikke kan endre lysfølsomheten lokalt i et punkt, fordi man da endrer integrasjonstiden, og det kan være problematisk, siden avlesning ved skanning, som er det vanlige, krever lik samplingstid for alle fotodetektorene. Ved å lage litt smartere fotokretser kan dette likevel gjøres, men på bekostning av noe areal.

4.5 Tilpasning til lyset

Tilpasning til lysintensiteten for å oppnå en stor dynamisk spennvidde kalles adaptasjon. Det er tidligere nevnt at fotostrømmen kan variere med over 6 dekadere. Siden kretsene vanligvis fungerer over en relativt liten spennvidde, må det gjøres en tilpasning så tidlig som mulig. Man snakker om to typer adaptasjon.

4.5.1 Global adaptasjon

Global tilpasning kan sammenliknes med iris i øyet. Iris er den fargede delen av øyet og har en åpning som kalles pupillen. Størrelsen på denne åpningen justeres for å øke eller minske lysmengden som slipper inn i øyet. For kameraer heter den tilsvarende delen blenden, som sitter i objektivet, og som også er en justerbar åpning for å justere den totale lysmengden. Felles for disse to er at det dreier seg om en global tilpasning for alle cellene i netthinnen og alle de lysfølsomme elementene i fotosensoren.

Selv om global adaptasjon vanligvis gjøres med blenden, kan den også gjøres elektronisk på sensoren, ved å justere forsterkningen felles for alle billedpunktene. Fordelen med dette er at man kan benytte et enklere objektiv uten blende. Ulempen er at man ikke får fordelen av høyere dybdeskarphet ved små blenderåpninger, men tvert i mot får den laveste dybdeskarpheten i alle lysforhold.

4.5.2 Lokal adaptasjon

Lokal tilpasning er mer avansert, og er en av ideene fra det menneskelige øye (Bohaen 1996). Med lokal adaptasjon justerer individuelle celler/billedpunkter seg for å oppnå en god forsterkning av mørke områder samtidig som lyse områder i andre billedpunkter dempes. For å se at dette er implementert i øyet, er det enkelt å ta et bilde med et helt vanlig kamera av en scene der det er stor forskjell på lyseste og mørkeste punkt, for eksempel et rom med vinduer. Da vil kameraet enten innstille seg til lysnivået ute (gjennom vinduet) eller inne. Øyet vårt vil derimot kunne vise detaljer både inne og ute, som er et tydelig eksempel på lokal adaptasjon.

Den beste tilpasningen av lyset får man med en kombinasjon av lokal og global adaptasjon. For å lage dette i mikroelektronikk kan man implementere adaptasjon på flere nivåer.

Første nivå er de lysfølsomme elementene, fotodetektorene. Man kan fremstille fotodetektorer som er lysadaptive. Ved å kople slik at svakt lys aktiverer større følsomhet enn sterkt lys, kan man forbedre dynamikken en god del, men ikke veldig mye. Det er likevel av stor verdi, siden dette er det signalet som arbeides videre med i fotokretsene. Støyforholdet holdes noenlunde konstant. Dette er en lokal adaptasjon, siden hver fotodetektor arbeider for å gjøre sin dynamikk så stor som mulig uavhengig av andre billedpunkter.

Fotokretsene er neste nivå. Her er det mer fleksibelt hva man vil gjøre enn i detektorene. Kretsene med en logaritmisk kompresjon av signalet har så absolutt en adaptasjon, selv om det minsker kontrastene. I integrasjonsbaserte fotokretser gjør man adaptasjon ved å kontrollere integrasjonstiden, som kan sammenliknes med lukkertiden på et kamera. Begge disse formene for adaptasjon er lokal.

Videre kan man kople kretsene til spesielle lysadaptive arkitekturer. Til forskjell fra fotokretsene er det mer systematisk/matematisk lagt opp, men ellers ingen særlig forskjell.

En mer avansert måte å tilpasse lysintensiteten er ved å sammenlikne strømmene med et lokalt gjennomsnitt. Da unngår man den store utjevningen av kantene som forekommer ved ren lokal adaptasjon. En del støy kan også fjernes.

Global adaptasjon styres på øverste nivå. Da ser man f.eks. på gjennomsnittet av alt lys og styrer f.eks. en blender eller lukkertiden etter dette.

4.5.3 Aktive sensorer

Fotokretser som forsøker å forbedre billedkvaliteten, kalles ofte aktive sensorer. De forsøker ikke å detektere bevegelse og andre operasjoner på høyt nivå, og sees derfor ofte på som mindre smarte sensorer. Fotokretsen i en aktiv sensor er ofte basert på "sample-and-hold" integrasjonsmodellen beskrevet tidligere, men dette er ikke et krav.

En av de store ulempene ved CMOS kameraer i forhold til CCD kameraer er forskjellen mellom transistorene som skal være identiske. Et hvitt ark blir derfor ikke hvitt ut fra kameraet, men mange forskjellige gråfarger. Slik støy kalles ofte "salt og pepper støy". En enkel metode å fjerne denne støyen på er å sample alle fotokretsene når fotodetektorene er slått av. Man får da målt forskjellen mellom transistorene. Denne forskjellen kan man så lagre og trekke fra senere for å utligne. Et problem imidlertid er at det tar svært stor plass å lagre denne informasjonen.

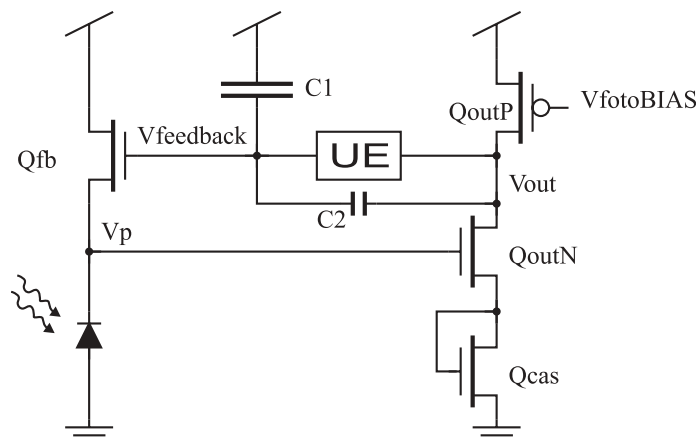
En måte å fjerne salt og pepper støyen på, uten å måtte lagre dette for hvert billedpunkt, er å bruke korrelert dobbel sampling. Det vil si at man deler hver tidsperiode i to like deler, der man i første halvdel samler med fotodetektoren slått av, og i andre halvperiode med fotodetektoren på. Man har da den informasjonen man trenger. En ulempe er selvfølgelig at man halverer integrasjonstiden som lyset slipper til, og dermed også lysfølsomheten. Helt fjernet blir støyen likevel ikke, siden transistorene som tar hånd om av og på signalene ikke er like.

Felles for disse fotokretsene er at de bygger på "sample-and-hold" konseptet, som egner seg best for skanning. De er derfor ikke aktuelle i kameraet.

4.6 Fotokretsen i kameraet

I likhet med det biologiske øyet, ønsket jeg at kameraet skulle tilpasse seg lyset lokalt i hver piksel, for å gi kameraet evnen til å kunne se i mange lysforhold, bl.a. se i skygger, se mot solen (eller vindu m. sollys) og liknende. Siden kameraet skal se etter endringer, er det viktig å få temporær informasjon tydelig frem. Statisk informasjon, som lysnivået, er mer eller mindre uvesentlig.

Fotokretsen er basert på Delbrück og Meads adaptive fotokrets (Delbrück et al. 1994) har disse egenskapene, og som også passer godt med valget av lysfølsomt element (kan ikke benyttes med integrerende fotodetektor). Denne fotokretsen er dessuten kompakt, som er viktig når den skal brukes i hver eneste piksel. I forhold til denne kretsen er størrelsesforholdet $C1/C2$ er dramatisk endret ($C2$ er meget liten, $C1$ er veldig stor) og en diodekoplet transistor, Q_{cas} , er lagt til for å heve V_{out} med ca. 0,7 V. Fotokretsen er vist i Figur 7.



Figur 7 - fotokretsen

Å forsterke fotospenningen byr på en stor utfordring, nemlig at fotostrømmen kan variere med så mye som 10^6 . Det gjør at hvis vi ønsker lineær forsterkning, vil vi ikke kunne forsterke særlig mye uten å begrense oss til et mindre område av lyset. En mulighet er nettopp å gjøre dette, begrense seg til et mindre område av lyset, men samtidig la dette området være justerbart. Det blir virkelig elegant når dette området justerer seg automatisk.

4.6.1 Analyse av fotokretsen

Generelt om kretsen kan det sies at den gjør store utslag for endringer innen det innstilte lysnivået. Endring av lysnivå, som vil bli en stor DC endring, tilpasses raskt gjennom det ulineære elementet, markert UE i figuren. Veldig raske endringer, som vil være AC strøm med høy frekvens, koples gjennom kondensatoren C2 og filtreres bort.

Dersom fotodioden eksponeres for en økning i lyset vil fotostrømmen øke, og dermed senke spenningen V_p . Denne spenningen går inn på den inverterende forsterkeren som består av transistorene Q_{outP} og Q_{outN} , og gjør at V_{out} gjør et stort positivt utslag.

Utgangsspenningen V_{out} koples nå tilbake gjennom C_2 og UE, avhengig av dens AC/DC karakteristik. Det ulineære elementet UE har en strøm/spenning karakteristik som er slik at for en spenningsforskjell over en viss størrelse, vil en stor strøm gå gjennom elementet, og dermed gi rask tilbakekopling. For små spenningsforskjeller vil nesten ingen strøm gå gjennom UE, og må tilbakekoples gjennom C_2 . Den tilbakekoplete spenningen lagres på kondensatoren C_1 , som er laget ekstra stor for å gjøre tilpasningen treg.

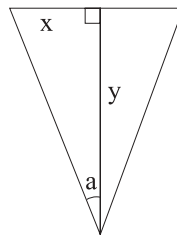
Når $V_{feedback}$ øker vil også strømmen gjennom Q_{fb} øke. Dette medfører at V_p igjen stiger til nesten samme nivå den hadde før endringen fant sted. Dess sterkere forsterkeren og tilbakekoplingen er, dess fastere holdes V_p .

Størrelsen av C_1 er med på å bestemme hurtigheten av tilbakekoplingen. For kameraets vedkommende er det ønskelig med en noe treg tilbakekopling, for å kunne detektere trege endringer. Strømmen gjennom UE bestemmer også hastigheten for endring av kondensatorens spenning (tidskonstant = RC), men det ulineære elementet kan ikke endres særlig på, selv om et alternativ ville være å bruke et annet ulineært element. Størrelsen av C_1 er bestemt ved simuleringer, og beskrevet i neste avsnitt.

4.6.2 Simuleringer av fotokretsen

Det er utført mange simuleringer under arbeidet med kretsen. For hver endring måtte nye simuleringer utføres. Etter at fotodetektoren, fotokretsen og endringsdetektoren var bestemt måtte bl.a. størrelsen på kondensatoren C_1 (Figur 7 side 28) bestemmes. Denne avgjør hvor trege endringer kameraet er i stand til å merke.

Kameraet har en båndpass filtrering i det det kun registrerer endringer i et visst frekvensområde. Ved å gjøre C_2 meget liten, er det kun veldig høye frekvenser som kuttes bort. Simuleringene fokuserte mest på nedre kuttfrekvens, som bestemmes av C_1 (bestemmes også av det ulineære elementet, UL i Figur 7, men dette kan ikke endres særlig). Nedre kuttfrekvens ble bestemt slik:

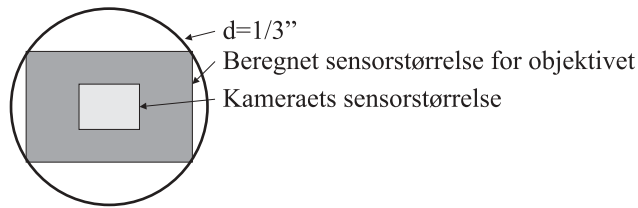


Figur 8 - hvor bredt område kameraet kan se

I Figur 8 er vinkelen til kameraet = $2a$. Avstanden vi ser (og fokuserer på) er y og kameraets synsfelt = $2x$. Trigonometri gir oss

$$x = y \cdot \tan(a) \quad (6)$$

Et bruksområde for kameraet vil være å stå og se utover et stort område, f.eks. en parkeringsplass. Bevegelser på parkeringsplassen, som biler og personer som kommer og går, detekteres av kameraet og kan eventuelt brukes til å styre et zoom kamera. En tenkt avstand fra kameraet til midt på plassen kan være 50 m.



Figuren viser en skisse av det forminskede bildet fra linsen. Vinkelen til linsen diagonalt er $76,9^\circ$ for et "footprint" på $1/3''$ (0,85 cm), som er omtrent fire ganger så stort som sensoren. Sensoren ser derfor vinkelen $76,9^\circ/4 \approx 19^\circ$. Vinkelen α i Figur 8 er halvparten, altså $9,5^\circ$. Synsfeltet til kameraet i meter blir da:

$$x = 50m \cdot \tan(9,5^\circ) \approx 8,4m \Rightarrow 2x = 16,8m \quad (7)$$

Jeg antar at kameraet vil ha rundt 50 piksler, og får

$$\frac{16,8m}{50 \text{ piksler}} = 0,34m/\text{piksel} \quad (8)$$

Hver piksel overvåker et område med bredden 0,34 m i virkeligheten. Hvis et objekt som beveger seg med 1 m/s skal kunne detekteres må nedre kuttfrekvens være:

$$\frac{0,34m/\text{piksel}}{1m/s} = 0,34 \text{ s}/\text{piksel} = 2,9Hz \quad (9)$$

Det ble utført simuleringer med stige- og falltid på 0,1 s, 0,5 s og 1 s, der de to første ble lagt mest vekt på. Størrelsen av kondensatoren C_I ble bestemt som følge av dette. Det kan være ønskelig å detektere bevegelser på 50 meters avstand som er tregere enn 1 m/s, men en avveining er gjort her for å unngå at C_I blir for stor.

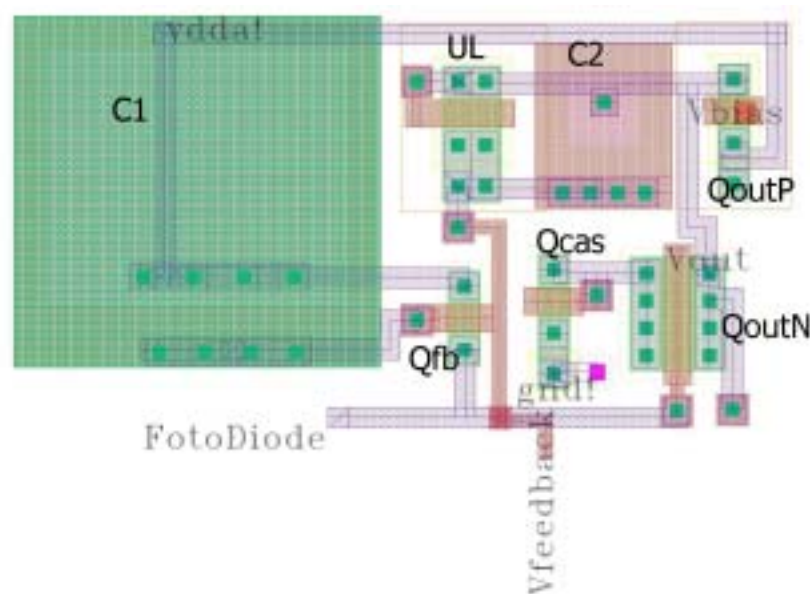
4.7 Utlegg av fotokrets

Utlegg av fotokretsen er vist i Figur 9. Krets-elementer fra Figur 7 er markert. I kameraet er fotokretsen dekket med metallag 3, for å skygge for lyset, men dette er ikke vist her. Dette metallaget brukes også til å distribuere *jord* på kretsen.

Som utlegget viser, er kondensatoren C_I svært stor sammenliknet med resten av elementene. Den er bygd opp av to polysilisium lag, og måler $15,5\mu m \times 16,1\mu m$ i ytterkant, hvilket gjør den større enn fotodioden. Arealet som går med er $\approx 250 \mu m^2$, men bare $147 \mu m^2$ utgjør kondensatoren (området med begge polysilisium lag) på grunn av at lag 1 må innramme lag 2 med minst $1,5 \mu m$. Dette gjør at en kvadratisk form er mest arealbeparende når man konstruerer kondensatorer av denne typen. Kapasitansen er på 132 fF. En ting man kan merke seg er at kryssing av metallag 1 og

2 ikke er lov over en slik kondensator. Siden den er såpass stor, og fotodioden også er stor, ble rutingen over matrisen av endringspikslers vanskeligere.

Videre er elementene som trenger en n-brønn plassert øverst og øvrige nederst. Det ulineære elementet trenger en egen brønn (som ikke er tilkoplest Vdd), og deler derfor ikke brønn med Q_{outP} .



Figur 9 - utlegg av fotokrets

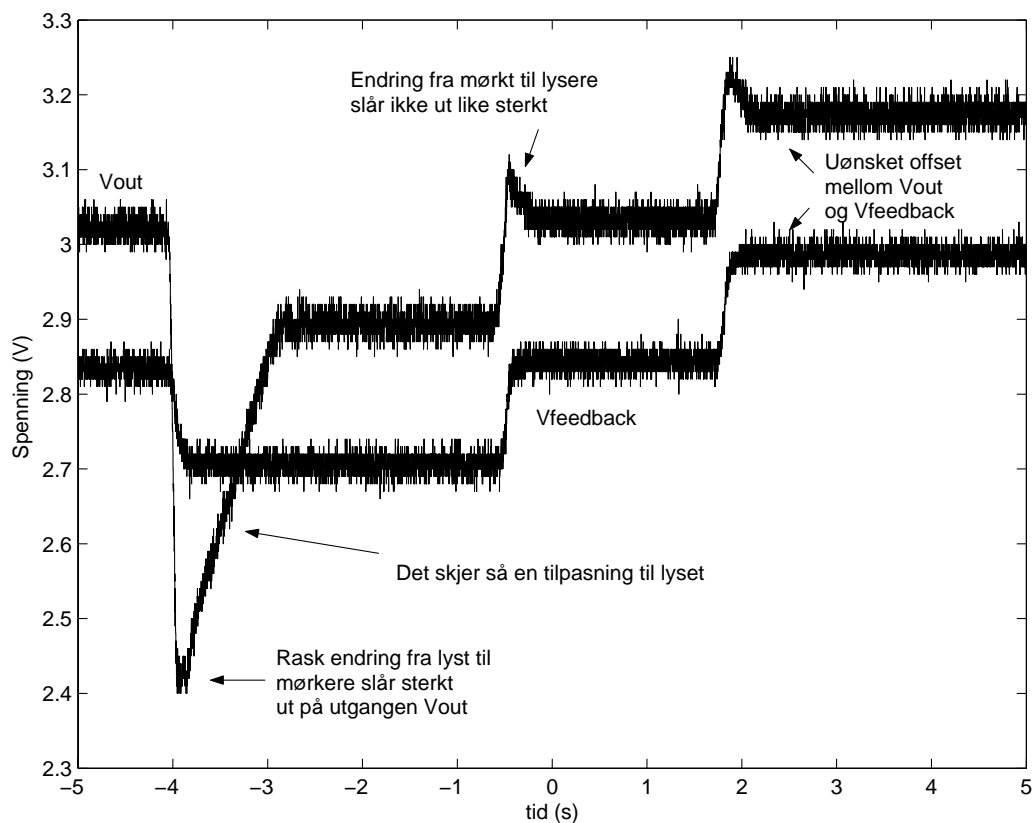
4.8 Målinger på fotokretsen

Utgangene som er målt er fotokretsens V_{out} og $V_{feedback}$, som er inngangene til korrelasjonskretsen. Fotokretsen er målt med flere fotoelementer:

- ▶ Fotodiode, nøyaktig som i hovedkretsen. Dette var selvfølgelig den viktigste av målingene.
- ▶ En stor nMOS transistor som erstatning for fotodioden. Kretsen ble da målt uten lys, og lyset ble simulert ved å endre på gate spenningen til transistoren, som hele tiden ble holdt i svak inversjon, for å etterlikne de små strømmene det går i en fotodiode (reversstrøm).
- ▶ Fotogate, som er nMOS transistoren fra forrige punkt, men denne gang eksponert for lys. Gate spenningen holdes fast, men kanalen i transistoren modelleres av lyset.

I alle tre tilfeller ble resultatene nokså like, selv om det er store forskjeller mellom elementene i listen over. Dette kommer av fotokretsens oppbygning med en tilbakekopling som er mye sterkere enn signalet selv. Så selv om fotogate og eventuelt bipolare transistorer har større strøm, så er det fotodiodens motstand mot støy som er viktigst når denne fotokretsen benyttes.

I Plott 1 ser vi fotokretsens store utslag for endringer og etterfølgende tilpasning til lyset. Siden det allerede er mange plott der fotodiode er benyttet (og kretsen er eksponert for lys), for eksempel øvre delplott i Plott 5 på side 45, er følgende et plott der hele kretsen er tildekket. Fotodioden er her byttet ut med en nMOS transistor, som forklart. Fordelen av å kunne teste kretsen uten lys er at man slipper å tenke på de bivirkninger som eksponering for lys har, da de fleste strukturene på brikken blir mer eller mindre påvirket av lys. Sammenliknet med øvre delplott i Plott 5 ser vi at oppførselen er tilnærmet lik. Nivåene ligger også noenlunde likt i dette plottet som i plottet der vi eksponerer for lys, kun en anelse høyere, men så er også kretsen kun eksponert for belysning av en liten lommelykt i Plott 5.



Plott 1 - testing av fotokretsen der fotodioden byttet ut med transistor (tester uten lys)

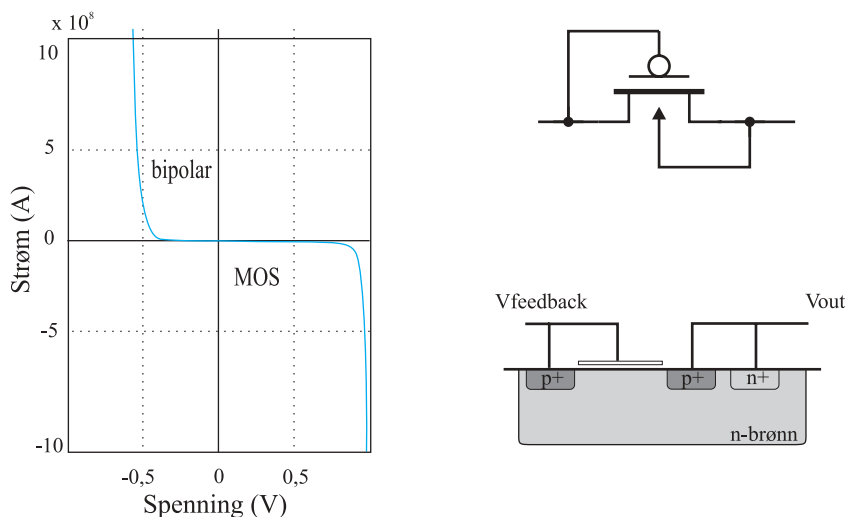
Den øvre kurven er V_{out} , som er en invertert forsterket utgave av diodestrømmen, den nedre kurven er $V_{feedback}$, som henger litt etter V_{out} i tid, og som ikke får med seg raske endringer. Dette fordi $V_{feedback}$ integreres på en kondensator, og derfor ser glattet ut i forhold til V_{out} . Plottet viser at fotokretsen tilpasser seg "lyset" (ser da på V_{out}).

4.8.1 Sammenlikning med simuleringene

En umiddelbar forskjell fra simuleringene er en uønsket avstand mellom kurvene på ca. 200 mV. Dette viste seg *ikke* på simuleringene, der kurvene i "hviletilstand" var nesten helt like. Denne avstanden lager store problemer for funksjonaliteten som

endringsdetektor, der nettopp en slik avstand tolkes som endringer, mens $V_{out} \approx V_{feedback}$ tolkes som at ingen ting skjer.

Etter at dette ble oppdaget på målinger, analyserte vi kretsen i nærmere detalj. Det viser seg at det ulineære elementet har en ganske stor lekkasje mot jord på den ene siden (brønn til substrat). Vanligvis vil denne siden stå koplet mot p-transistoren som fungerer som en strømkilde. Denne transistoren vil da levere nok strøm til at lekkasjen ikke gjør noe fra eller til, og det er slik det ulineære elementet er ment koplet opp. Hvis man ikke ser nøye etter, kan man fort gjøre den feilen jeg har gjort; å anta at det ulineære elementet er symmetrisk, og kople det opp i feil retning. Lekkasjen kommer mot kondensatoren C_I , og dette gjør at $V_{feedback}$ hele tiden ligger lavere enn V_{out} . Det ulineære elementet er vist i Figur 10 i den retningen det burde vært koplet opp i fotokretsen.



Figur 10 - ulineært element, I/V kurve, fysisk skisse og skjema.

Simuleringer bekrefter at denne lekkasjen er grunnen til avstanden mellom kurvene. I nye simuleringer er en substrat-diode (brønn til substrat) lagt til for å eksplisitt legge inn en lekkasje. Når det ulineære elementet med lekkasje diode står samme retning som på brikken, får vi den uønskede avstanden mellom kurvene V_{out} og $V_{feedback}$ som vi observerer på målinger. Ved å snu elementet riktig vei, slik at lekkasjen tar strøm fra p-transistoren Q_{outP} (se Figur 7) som står som strømkilde, forsvinner avstanden mellom V_{out} og $V_{feedback}$, så det er sannsynlig at dette er problemet.

4.8.2 Kort analyse av grafen

Ved tiden $t \approx -4$ sekunder skjer det en rask endring fra lyst til mørkere, som så holdes stabil frem til $t \approx -1$ s. Utgangen V_{out} gjør et stort utslag når endringen finner sted, fra over 3V til ca. 2,4 V. Over en periode på omtrent et sekund tilpasser kretsen seg så det nye (mørkere) lysnivået, som kan sees på utgangen som går mot et nytt spenningsnivå på 2,9 V, som er nesten likt spenningsnivået den hadde (litt over 3 V).

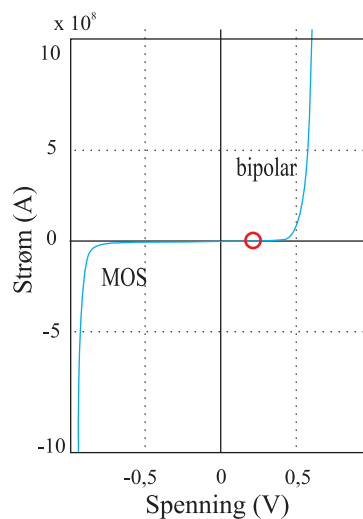
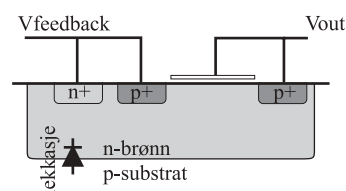
Den originale adaptive fotokretsen er modifisert, slik at den skal tilpasse seg så sent som mulig, hovedsaklig ved å gjøre ”lære-kondensatoren” C_1 stor og kondensatoren C_2 i tilbakekoplingen veldig liten. Sistnevnte tvinger også raske endringer til å gå gjennom det ulineære elementet. Dermed bruker kretsen et helt sekund på å tilpasse seg, som er det vi ønsker.

Etter at kretsen har stabilisert seg, gjøres så en ny endring ved $t \approx -0,5$ s, denne gangen fra mørkt til litt lysere. Igjen observeres et stort utslag til ca. 3,1 V, som så gradvis modererer seg til 3 V. Tilpasningen i denne retning, altså fra mørkt til lysere, går mye raskere enn motsatt. Dette ser vi også på Vfeedback, som endrer seg raskere, og dermed gjør at utslaget ikke blir like stort som for endringer andre veien.

4.8.3 Ulik tilpasningshastighet opp og ned

Det kan være flere grunner til at endringer opp og ned tilpasser seg med ulik hastighet. For det første er det ulineære elementet (som tilbakekoplingen går gjennom) asymmetrisk, fordi det fungerer enten som en bipolar diode eller en MOS diode, avhengig av retningen. Dette fører til at lære-kondensatoren C_1 (Figur 7) sin oppladningsstrøm og utladningsstrøm, som går hver sin retning gjennom det ulineære elementet, ikke er like store. Lære-kondensatoren lades opp raskere enn den lades ut.

For $V_{out} > V_{feedback}$ vil vanligvis MOS transistoren være på og bipolar være av. Siden jeg har koplet omvendt vil bipolar være på. Nå som $V_{out} > V_{feedback}$ i stabil tilstand, vil stabil tilstand se slik ut:



Figur 11 - ulineært element slik jeg har koplet det opp i fotokretsen. Lekkasje markert i fysisk skisse. Spennning markert på I/V kurven for stabil tilstand (ca. 0,2 V). Endringer til lysere vil bevege oss mot høyre i kurven og gi en stor strøm.

Figur 11 viser hvordan det ulineære elementet er koplet opp i kameraet, altså feil vei. Lekkasjen er markert med diodesymbol (i sperreretning), og er betydelig på grunn av diodens størrelse (størrelsen av brønnen).

Den andre grunnen til at fotokretsen tilpasser seg endringer til mørkere tregere enn endringer til lysere har med lekkasjen å gjøre; at $V_{feedback} < V_{out}$. Som vist i Figur 11

over, gjør dette at det ulineære elementet (UE) befinner seg i tilstanden markert med en sirkel, når ingen endringer skjer.

Ved en endring til lysere, fører en liten økning av V_{out} til en stor strøm gjennom UE og umiddelbar tilpasning til lyset, fordi $C1$ lades raskt opp. I Figur 11 vil vi bevege oss mot høyre i grafen, som gir en stor strøm gjennom elementet.

Ved en endring til mørkere, vil strømmen først være svært liten når vi går mot venstre i kurven. Først når V_{out} blir mye mindre enn $V_{feedback}$ vil strømmen bli av en størrelse som gjør at $C1$ lades raskt ut. Derfor går tilpasning til mørkere tregere enn motsatt.

En kommentar til dette vil være at spenningen på $C1$ vil synke allerede før $V_{feedback} > V_{out}$, pga. lekkasjen. Så selv om det går strøm mot venstre i UE vil netto strøm til $C1$ være negativ, og spenningen vil synke. Dette vil likevel være lite, og tilpasningen blir treg.

Til slutt kan det nevnes at det er en del støy på målingene som vi kan se. Dette er naturlig støy, bl.a. termisk støy, og ikke fra lys, siden Plott 1 viser målinger med lokk på kretsen.

4.9 Om testing med lys

Testing med lys er vanskeligere å kontrollere enn tester basert på spenning. Det er også vanskeligere å dokumentere. Med spenning kan man måle nøyaktig hva inngangssignalet er, og bruke signalgeneratorer for å automatisk stimulere kretsen. Med lys som skal bevege seg, vil dette gjøres manuelt. Og uten en lysmåler, er det vanskelig å måle hvor stor lysmengden er.

En annen utfordring når det gjelder kretser som er følsomme for endringer i lyset, er forstyrrelser. Da jeg først målte utgangen fra fotokretsen fikk jeg store forstyrrelser, og trodde at fotodiodene ikke virket som forventet. Ved å øke oppløsningen på oscilloskopet kunne jeg se at forstyrrelsene var jevne halvsirkeler, som en likerettet sinus, med en frekvens på nøyaktig 100 Hz. Dette tilsvarer en sinus kurve med frekvensen 50 Hz. Det er da greit å forstå at disse forstyrrelsene må komme fra taklyset, som er tilkoplek strømnettet som har en vekselstrøm med frekvensen 50 Hz.

Testene måtte dermed utføres med alt lys slått av, slik at kun lys utenfra ga laboratoriet en svak belysning. Ved å endre på fotokretsen slik at forholdet C_1/C_2 har riktig størrelse kan frekvenser ≥ 50 Hz kuttes bort, som vil være nyttig for kretser som skal brukes/testes innendørs.

5 KORRELASJONSKRETS OG ENDRINGSDETEKTOR

Deteksjon av endringer er en av de mest fundamentale funksjoner ved syn, og sensorene er utviklet på grunnlag av dette. Faktisk er det slik at når vi fester blikket på noe, så er det mikroskopiske vibrasjoner i yet som gjør at sensorene blir stimulert selv om det vi ser på står stille. Fra i The human vision system funnet på nettstedet Wayfinding.net i hjemmesiden til Institute for Innovative Blind Navigation.

5.1 Deteksjon av endringer

Deteksjon av endringer er en temporær beregning. Slike beregninger krever informasjon om en tilstand tilbake i tid, som forklart i avsnitt 2.7. Dette involverer vanligvis å huske en tidligere verdi, men for deteksjon av endringer holder det å ha en kopi av tilstanden som forsinkes, og sammenliknes med tilstanden nå. En forsinkelse er innført ved den store kapasitive noden C_I i fotokretsen, vist i Figur 7 på side 28, som er nærmere beskrevet i avsnittene 4.6.2 og 4.7.

Til å sammenlikne de to spenningene V_{out} og $V_{feedback}$ fra fotokretsen behøves en korrelasjonskrets. En antikorrelasjonskrets ville vært bedre, siden vi ser etter når V_{out} og $V_{feedback}$ er *ulike*, men den valgte korrelasjonskretsen har såpass gode egenskaper i forhold til de vurderte alternativene, at den likevel ble valgt. Et alternativ som ble særlig vurdert var differensialpar forsterker, men denne krevde bl.a. likeretting av utgangen, og den hadde ikke noe større område som ville gi ≈ 0 utgangsstrøm.

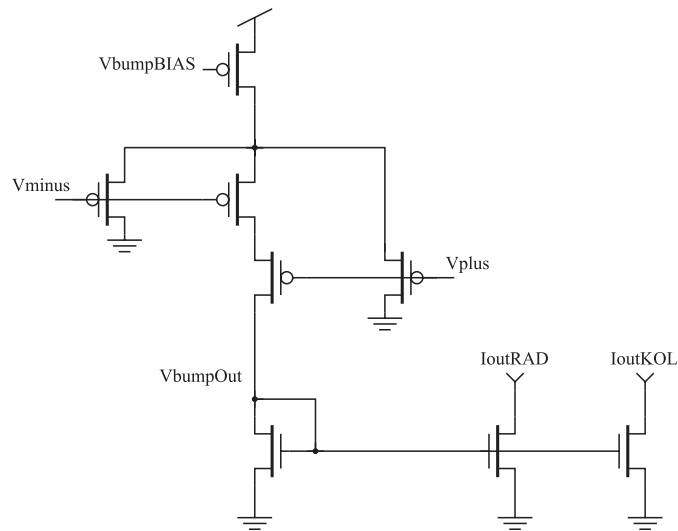
5.2 Beskrivelse av korrelasjonskretsen

Korrelasjonskretsen er basert Delbrücks ”bump-anti-bump” krets (Delbrück 1991). I forhold til den originale kretsen er den invertert, og et strømspeil er lagt til på utgangen (som inverterer tilbake og gir to utganger fremfor en). Den er meget kompakt, som er viktig fordi den brukes i hvert eneste fotopunkt – 2142 til sammen. Det er ønskelig at størst mulig areal i matrisen skal bestå av fotodioder, og minst mulig til andre kretser rundt disse. I tillegg har denne korrelasjonskretsen den egenskapen at den har en viss bredde på området $\Delta V (= |V_{out} - V_{feedback}|)$ der den gir utslag. Mange av alternativene er for følsomme, slik at selv små ΔV oppfattes som ulike. Siden V_{out} og $V_{feedback}$ aldri er helt like, selv når det ikke er noen endringer, er det viktig å ha en viss bredde på området ΔV .

”Bump-anti-bump” kretsen, som den også kalles, tar utgangspunkt i et vanlig differensialpar. I et vanlig differensialpar er strømmen gjennom halen gitt ved

$$I_b = I^+ + I^- \quad (10)$$

Når strømmen i den ene grenen øker må strømmen i den andre minke tilsvarende. Siden strømmen gjennom en transistor er eksponentielt styrt av spenningen i svak inversjon og kvadratisk i sterk inversjon, trenger ikke forskjellen i *gate* spenningene V_{plus} og V_{minus} være særlig stor for at den ene transistoren trekker mange ganger så mye strøm som den andre. Spenningen på den felles noden vil da synke, slik at drain-source spenningene for transistorene gjør at hele bias-strømmen går gjennom den ene grenen. Strømmen gjennom den andre transistoren vil være tilnærmet lik null.



Figur 12 - korrelasjonskretsen, "bump-anti-bump" kretsen, snudd og med et strømspeil på utgangen. Innganger er V_{minus} og V_{plus} , utganger er I_{outRAD} og I_{outKOL} for tilkopling til en rad og en kolonne.

Korrelasjonskretsen har to transistorer i tillegg til differensialparet. De to transistorene er koplet i serie, slik at strømmen gjennom denne tredje grenen begrenses av den transistoren med minst strøm. Når V_{plus} og V_{minus} er nær hverandre, vil bias-strømmen være gitt ved

$$I_b = I^+ + I^- + I_{midten} \quad (11)$$

Når V_{plus} og V_{minus} er ulike, vil strømmen i den seriekoblede midtre grenen også begrenses, slik at hele halestrømmen går gjennom enten høyre eller venstre gren, som for et vanlig differensialpar. Dette krever en relativ forskjell mellom V_{plus} og V_{minus} på ca. 250 mV. Oppsummert vil korrelasjonskretsen trekke strøm ved $\Delta V \approx 0V \pm 250 \text{ mV}$, og slutte å trekke strøm når $\Delta V > 250 \text{ mV}$.

Utgangsstrømmen blir så kopiert til I_{outRAD} og I_{outKOL} ved egenskapene til strømspeil. Strømspeil gir en god (invertert) kopi av strømmen dersom transistorene er i metning og de har samme *source* spenning. Ved at alle transistorene står mot jord, vet vi at

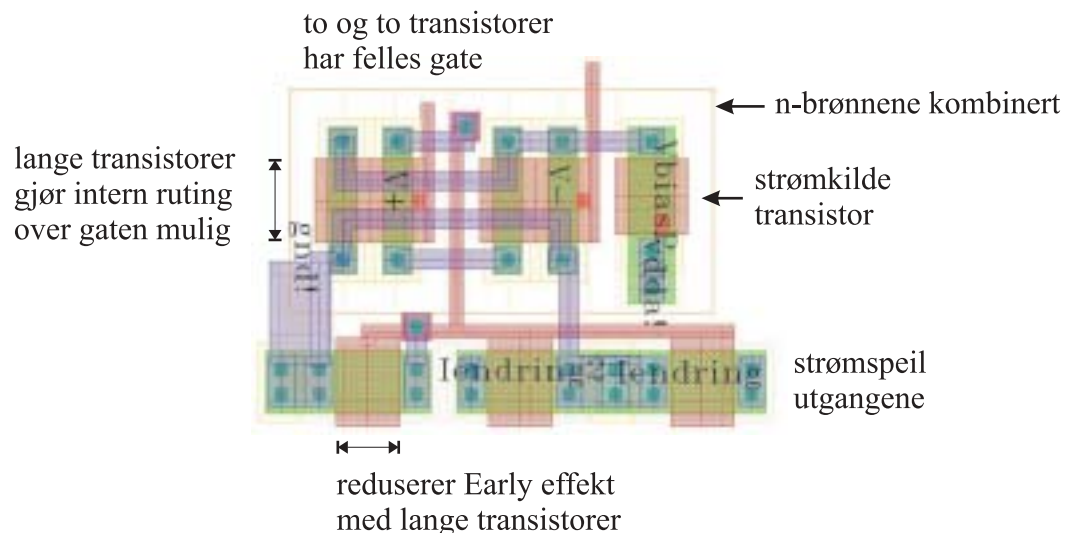
source spenningen er lik. Strømspeil er nesten uavhengig av *drain* spenningen så lenge vi har metning. For nøyaktige kopier kan store forskjeller i *drain* spenning gjøre Early effekten merkbar. I slike tilfeller kan man enten redusere Early effekten ved å lage lange transistorer, i kameraet er $3 \times 3 \mu\text{m}$ brukt, eller forskjellen i *drain* spenning kan reduseres ved å innføre en *cascode* transistor (Serrano et al. 1994).

5.3 Utlegg av korrelasjonskretsen

Utleppet av korrelasjonskretsen er meget kompakt. Som vist i Figur 13 er det flere ting som gjør dette mulig. For det første er n-brønnene for de fem øverste transistorene kombinert (se også Figur 12), som gjør at vi slipper å ha den store minsteavstanden mellom brønnene. For det andre er to og to transistorer koplet sammen med en felles gate, siden de har samme inngang. Lange transistorer gjør i tillegg til å forbedre relativ likhet i størrelser (transistor "mismatch") også at vi kan gjøre intern ruting over gatene. Til slutt er brønnskoplingene (til Vdd) satt inntil strømkildetransistoren istedenfor en separat kontakt. Strømspeil utgangene deler en felles substratkontakt.

Alle transistorene er større enn minimumstransistorer. $W/L = 2,2/3,6$ for "bump-anti-bump" delen (de fem øvre), og $3/3$ for strømspeil transistorene. For sistnevnte er det en fordel at transistorene er lange for å redusere Early-effekten, som kan gjøre speilingen dårligere. Minimumstransistorer i prosessen er $1/0,6$ og alle størrelser er i μm .

Til slutt kan det nevnes at korrelasjonskretsen skal passe sammen med fotokretsen, og at det skal være mulig å rute ledninger over den. Kun polysilisium og metallag 1 er benyttet. Total størrelse er ca. $16 \mu\text{m} \times 24 \mu\text{m}$.



Figur 13 - utlegg av korrelasjonskretsen

5.4 Målinger av korrelasjonskretsen

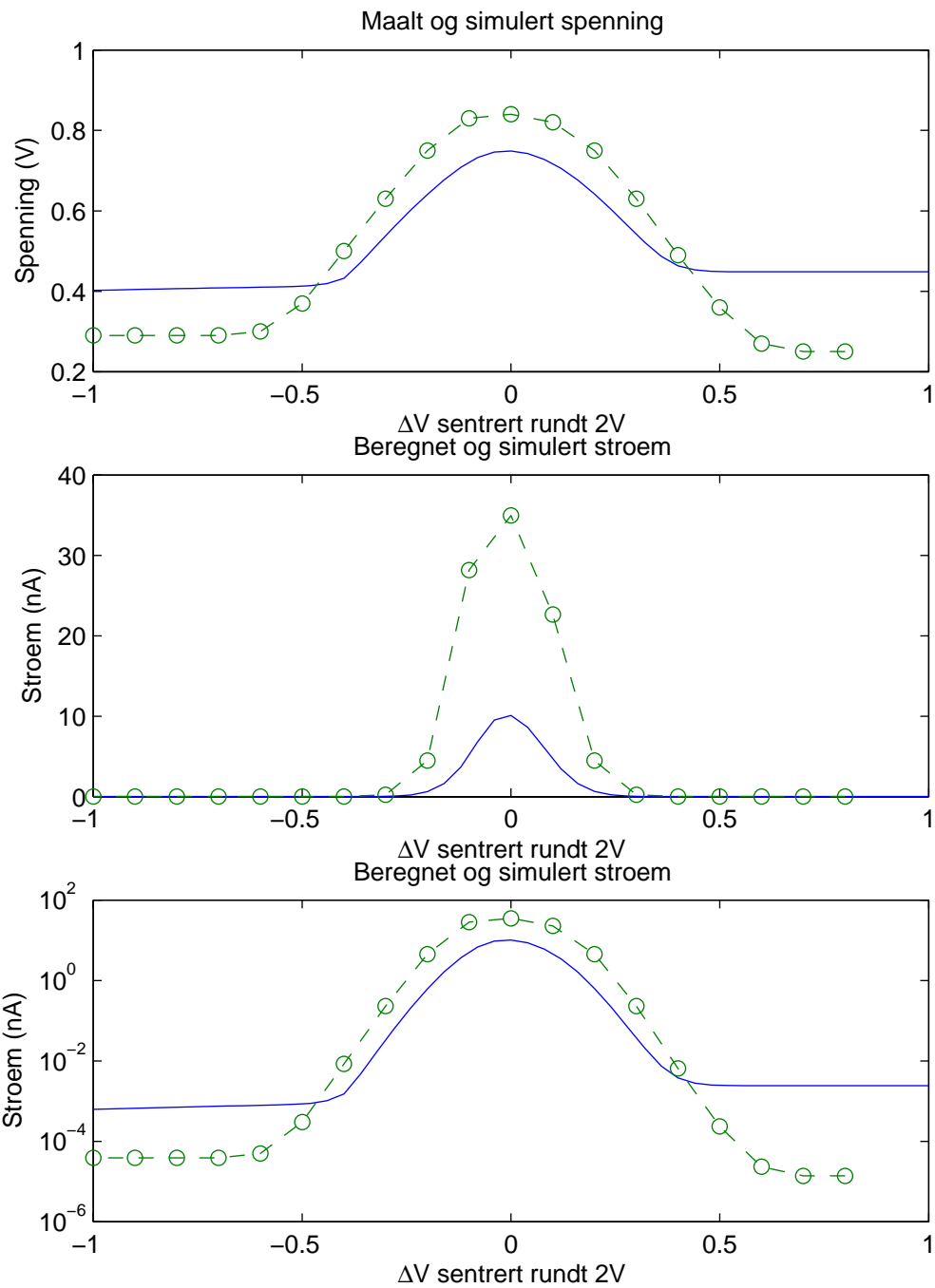
Utgangssignalet fra korrelasjonskretsen er vanskelig å måle, fordi det er strøm, og særlig fordi strømmene er veldig små (nA). Jeg har derfor valgt å måle spenningen $V_{bumpOut}$ over den diodekoblede n-transistoren, altså med *gate* og *drain* terminalene koplet sammen (se Figur 12). Denne koplingen medfører at $V_{gate-source}$ og $V_{drain-source}$ er like, og siden *source* terminalen er tilkoplek jord, holder det med et målepunkt for å kjenne både $V_{gate-source}$ og $V_{drain-source}$. Noden som måles er tilkoplek en spenningsfølger for å forstyrre signalet minst mulig, og videre tilkoplek en større spenningsfølger i pad-en som kan drive signalet ut av brikken.

I forhold til de andre kretsene er korrelasjonskretsen enkel å karakterisere, fordi den sammenlikner to faste spenninger. Prosedyren for måling var å sette den ene inngangen fast og variere den andre. For hver innstilling ble spenningen på utgangen målt, og resultatene ble ført opp i en tabell og vist grafisk i plottene nedenfor.

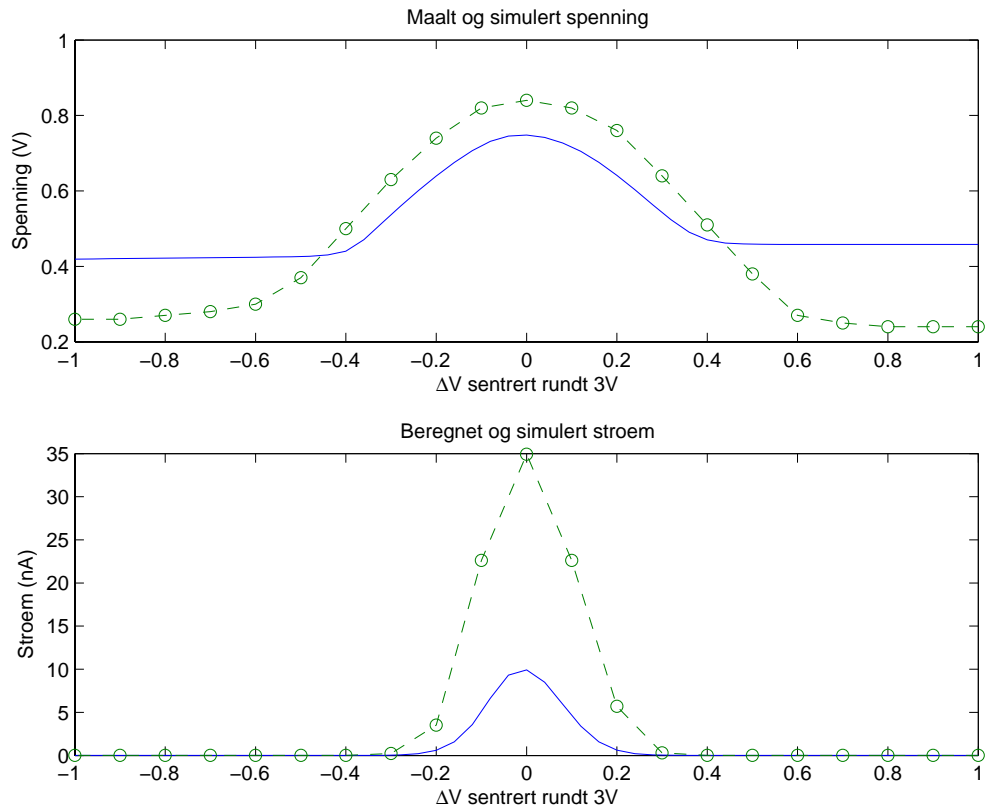
Simuleringer og målinger viser at utgangene V_{out} og $V_{feedback}$ fra fotokretsen, som er inngangene til korrelasjonskretsen når den brukes i hovedkretsen, ligger i området 2,4 til 3,5 V. For å undersøke kretsens virkemåte i hele arbeidsområdet, ble to sett målinger gjort, der den ene inngangen stod fast på henholdsvis 2 V og 3 V mens den andre varierte. Målingene ble så gjentatt på en annen brikke, begge ganger med lokket på for å unngå forstyrrelser lys kan gjøre på kretsene.

5.4.1 Resultater

Resultat fra målinger og simuleringer er kombinert i samme plott vist nedenfor for to ulike spenningsnivåer, 2V og 3V. Den interessante kurven i begge tilfeller, er den som viser strøm, siden det er dette som er utgangen av kretsen. Plott av spenning er tatt med siden det er dette jeg faktisk måler, og det blir lettere å sammenlikne simulering og måleresultat med større sikkerhet enn for strøm plottene.



Plott 2 - Målt vs. simulert spenning og beregnet vs. simulert strøm. En inngang holdes fast på 2 V mens den andre varierer mellom 1 og 3 V. Kurven med sirkeler er den målte. Nedre plott har logaritmisk akse.



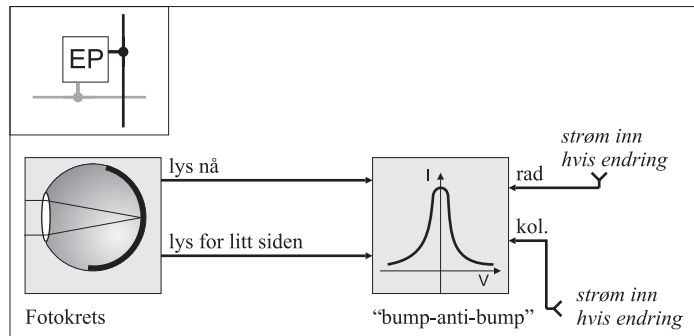
Plott 3 - Målt/beregnet vs. simulert spenning og strøm.
Den ene inngangen holdes fast på 3 V mens den andre varierer mellom 2 og 4 V. Kurven med sirkeler er den målte.

Formelen som er brukt i utregningen av strøm er EKV formelen (1) og (2) beskrevet i avsnitt 2.1. Strømmen er utregnet for venstresiden i strømspeilet.

Målingene viser at det er ganske bra samsvar mellom simuleringene og virkeligheten, med kun normale avvik. Kurveformen er lik for simuleringer og målinger. Bredden på kurven er ca. 250 mV til hver side (totalt 500 mV) som i beskrivelsen av kretsen. En ting å merke seg er at strømmen er lik for de to spenningsnivåene 2 V og 3 V (Plott 2 og Plott 3). Dette er fordi halestrømmen kontrolleres av en transistor som fungerer som en strømkilde med spenningen $V_{bumpBIAS}$.

5.5 Konstruksjon av endringsdetektoren

Det er nå enkelt å kople sammen fotokretsen med korrelasjonskretsen for å få en krets som ser etter endringer i lyset. I fotokretsen har vi spenningen $V_{feedback}$ som tilsvarer lyset for et øyeblikk siden $(t-1)$ lagret på kondensatoren C_I . Vi har også spenningen som tilsvarer lyset i dette øyeblikk (t) som V_{out} .



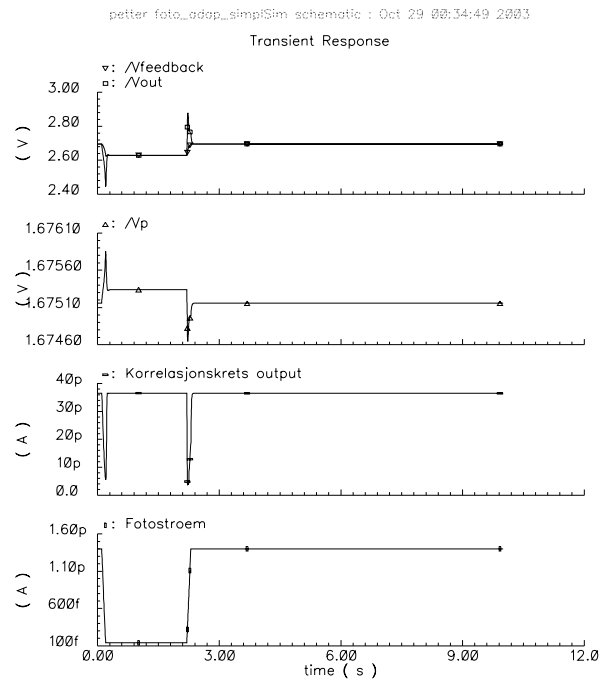
Figur 14 - en endringspiksel (endringsdetektor) er satt sammen av en fotokrets og en "bump-anti-bump" antikorrelasjonskrets.

$V_{feedback}$ og V_{out} koples inn på terminalene V_{plus} og V_{minus} på korrelasjonskretsen. Vi får da en krets jeg kaller *endringspiksel* som trekker strøm når lyset er stabilt, og slutter å trekke strøm ved endringer i lyset.

Endringsdetektoren er vist skjematisk i Figur 14. Det innfelte området viser symbolet brukt i Figur 4 som viser endringspikslenes rolle i hele kretsen.

5.6 Simuleringer

I simuleringer virket endringsdetektoren meget bra. Dette er vist i Plott 4, som er delt inn i fire deler. Øvre delplott viser V_{out} og $V_{feedback}$ i fotokretsen. Neste kurve viser spenningen over fotodioden, V_p , som holdes fast. Så vises utgangen fra korrelasjonskretsen, som går fra å trekke nesten 40 pA til nesten 0 får endringer både til lysere og mørkere. Nedre kurve viser fotostrømmen brukt til å simulere mørkeste nivå (flere nivåer er simulert).



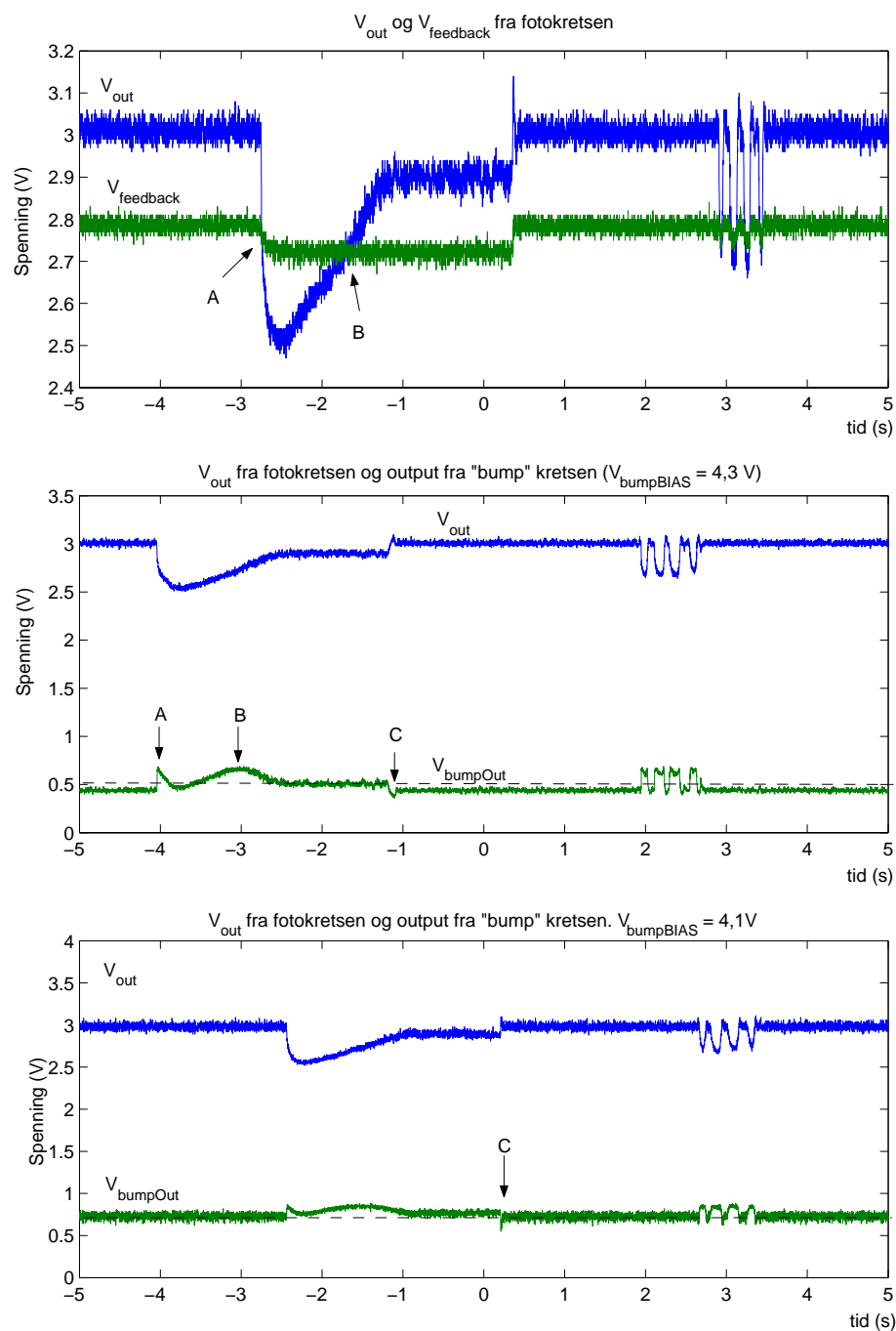
Plott 4 - simulering av endringspiksel.

5.7 Utlegg

Utlegg av endringspiksel er vist i Figur 15. Det er en fordel om hver celle er kvadratisk. Hvis ikke vil bildet være strukket i en retning, og bevegelser i ulik retning virke som de har ulik hastighet, selv om de er like. Når flere celler settes sammen blir arealet kvadratisk, og pr. piksel $36,8 \mu\text{m} \times 36,8 \mu\text{m} = 1354,2 \mu\text{m}^2$. Det vil si at vi har en dekningsgrad av lysfølsomme elementer på:

$$D = \frac{A_{\text{fotodiode}}}{A_{\text{total}}} = \frac{194 \mu\text{m}}{1354 \mu\text{m}} = 14,3\% \quad (12)$$

Figuren viser også at endringspikslene er dekket over med metallag 3, bortsett fra fotodioden som skal eksponeres for lys.



Plott 5 - V_{out} fra fotokretsen og $V_{bumpOut}$ fra antikorrelasjonskretsen for tre typer stimuli.

Det er verdt å merke seg at når det blir lysere, så går output fra korrelasjonskretsen ($V_{bumpOut}$) faktisk ned. Dette er markert ved bokstav C og er ønsket oppførsel. Ved endring til mørkere, som ved bokstav A, går output opp, som ikke er slik vi ønsker. Dette er pga. den uønskede avstanden mellom V_{out} og $V_{feedback}$ som kan sees i øvre

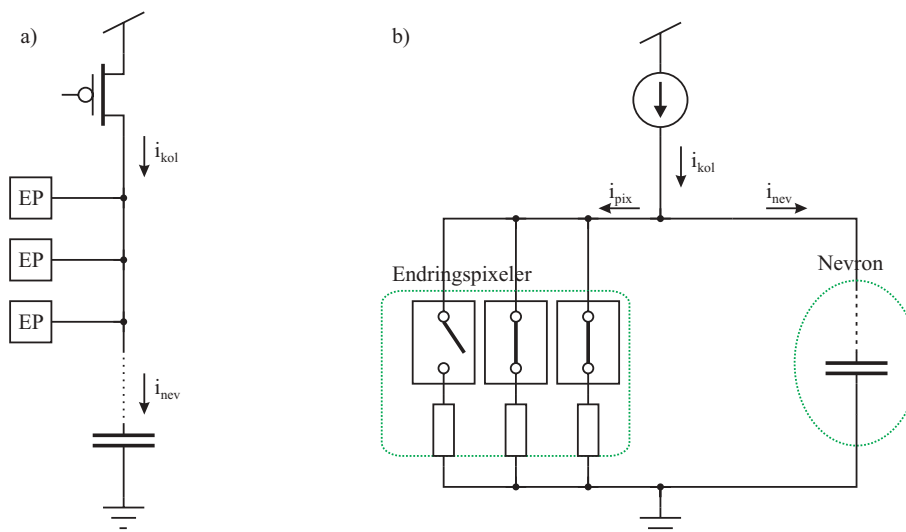
plott, og grundig analysert tidligere. Aktiviteten til høyre i plottene er fire fingre som beveger seg over linsen (tommelen ikke tatt med).

I det øvre delplottet markerer A og B der utgangene V_{out} og $V_{feedback}$ krysses. I disse punktene vil $\Delta V = 0$ for antikorrelasjonskretsen, som var ment å bety uendret tilstand. Utgangen fra bump kretsen vil befinne seg på toppen av kurven – den trekker maksimalt med strøm. Dette er markert av A og B i midtre plott, som viser gate og drain spenningene for strømspeilet på utgangen av antikorrelasjonskretsen.

Punkt C markerer utgangen fra endringspixelen ved en overgang fra mørkt til lyst. Det er her vi beveger oss utenfor ”bump” området, og kretsen slutter å trekke strøm, som betyr at endringer er detektert.

5.8.1 Sammenkopling av flere endringspikslers i en rad/kolonne

Dess mindre strøm endringspikslene i en rad eller kolonne trekker, dess mer strøm blir til overs til nevronene i winner-take-all kretsene. I figur a) i figuren nedenfor er dette vist. Strømmen for en rad/kolonne stilles inn med en bias spenning på p-transistoren øverst. Jo mer strøm endringspikslene (EP) trekker, jo mindre strøm får nevronet, som er vist som en kondensator som må lades opp. Meningen er at endringspikslene skal slutte å trekke strøm når de observerer endringer, slik at raden og kolonnen med mest endringer vil ha størst i_{nev} og bli valgt ut av winner-take-all kretsen. Figur b) viser oss at i_{nev} aldri vil bli 0 selv om alle endringspikslene trekker strøm, pga. parallellkoplingen.



Figur 16 - hvordan endringspikslene er sammenkoplet i en rad/kolonne

5.8.2 Innstilling av rad- og kolonnestrømmene

Som tidligere forklart, må kretsen stilles inn. Dette skjer bl.a. ved at i_{kol} i figuren over settes slik at i_{nev} blir liten nok til at ”timeout nevronet” kansellerer ut eventuelle pulser som i_{nev} kan føre til. Dersom det skjer endringer, og EP trekker

mindre strøm, skal inev bli stor nok til å generere pulser. Det gjelder derfor å finne et nivå slik at endringer vil være lavere enn nivået (trekke mindre strøm).

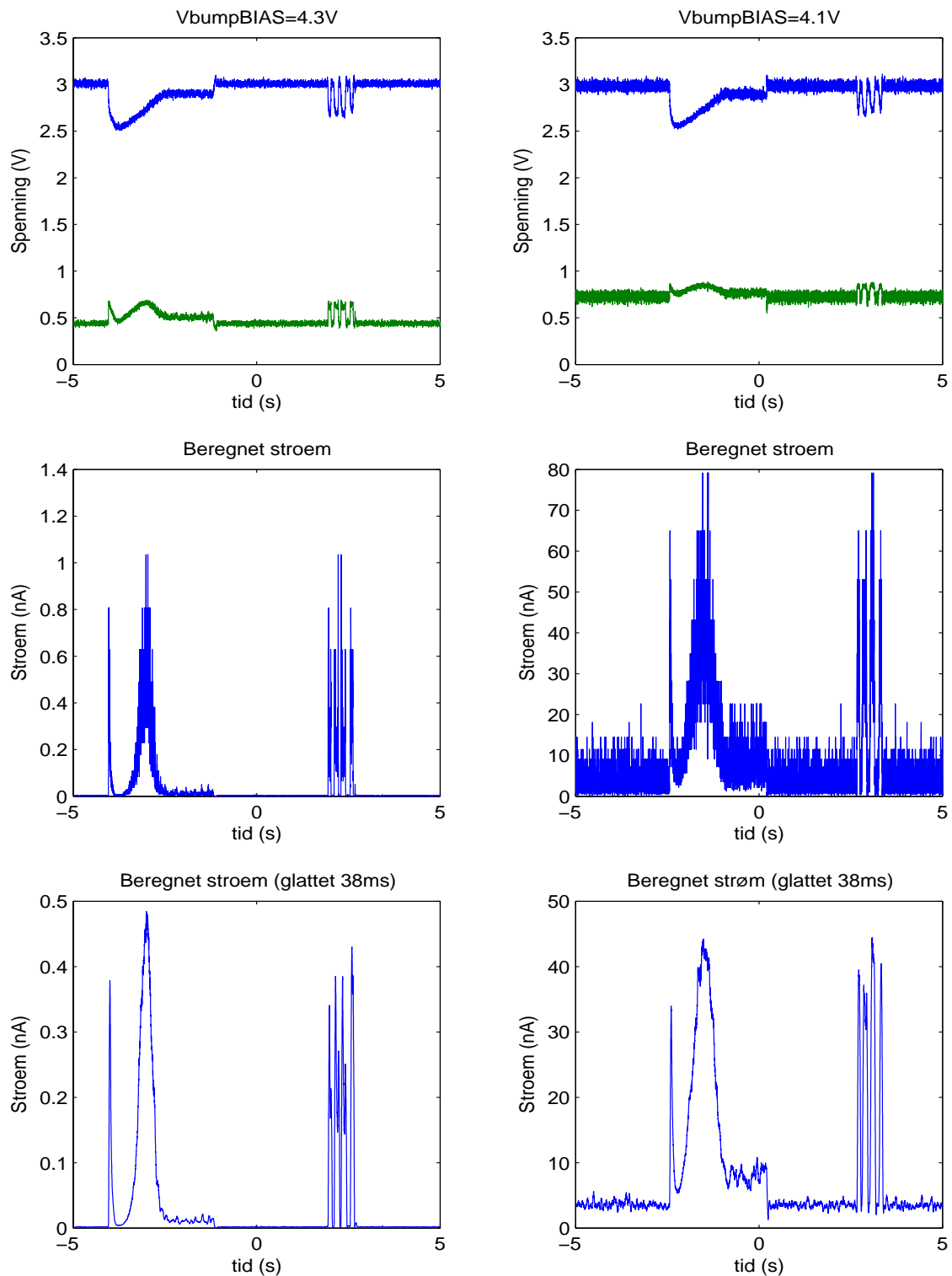
Den stiplede linjen i midtre delplott i Plott 5 viser stabilt nivå for totalt mørke. Vi ser at dette er litt høyere enn for stabilt lysere nivå. Dette skyldes at differansen mellom V_{out} og $V_{feedback}$ er større ved lys enn ved mørke, som kan sees i det øvre plottet. Stiller vi inn rad- og kolonnestrømmene til dette nivået (til å levere den strømmen dette nivået tilsvarer pr. piksel), vil det gå greit for mørke, mens økt (stabilt) lys vil gjøre at hver EP trekker for lite strøm, inev blir for stor, og vi får pulser selv om vi ikke har endringer.

Den stiplede linjen i nedre plott viser stabilt nivå (pr. piksel) for litt mer lys. Stiller vi inn rad- og kolonnestrømmene til å passe med dette nivået, vil vi ikke få pulser for verken stabilt lys eller mørke (ønsker jo ikke pulser ved stabilt nivå). Ved endringer ønsker vi pulser, som vi får dersom en del av pikslene kommer under nivået som den stiplede linjen markerer. Som vi ser av plottet skjer dette bare for endringer mot lysere, markert ved bokstav C.

På laboratoriet er dette testet, men det viste seg å ikke være mulig å få til, da støyen er nesten like stor som utslaget ved bokstav C. Mer nøyaktige innstillinger av kretsen kunne gjort dette mulig, men jeg har valgt å ikke bruke mer tid på å få dette til.

Nivået stilles derfor som i midtre delplott. Dette vil føre til at ikke bare endringer detekteres, men at alt lys oppfattes som endringer. Globalt sett vil kretsen finne lyseste punkt og følge dette. Dersom punktet stopper å bevege seg vil kretsen fremdeles se det, selv om det ikke endrer (flytter) seg, og "timeout" er det vi egentlig ønsket. Dersom lyspunktet forsvinner ut av synsfeltet til kameraet, slik at kameraet bare ser mørkere ting enn det innstilte nivået, vil vi få "timeout".

Testene i Plott 5, for både $V_{bumpBIAS} = 4,1 \text{ V}$ og $4,3 \text{ V}$, er vist i Plott 6 nedenfor sammen med beregnet strøm. Formlene som er brukt er (1) og (2) fra EKV modellen beskrevet tidligere.



Plott 6 - Strøm inn til endringspiksel. Øvre kurve viser utgangen V_{out} fra fotokretsen og spenningen på utgangen av korrelasjonskretsen. Midtre plott viser beregnet strøm med Vittoz EKV likninger. Nedre plott er glattet (gjennomsnitt av 38ms med målinger) for å se gjennom støyen.

6 NEVRONER OG WINNER-TAKE-ALL KRETSEN

6.1 Nevromorf koding

Hjernen består av nevroner og koplinger mellom dem. Ut fra og inn til nevronet går det mange tråder som kalles axoner. Knutepunktene på disse trådene kalles synapser. En synapse fungerer som en transistor. Den åpner og lukker for gjennomstrømming av ioner. Nettverket av koplinger mellom nevronene er derfor svært komplisert.

Nevronene kommuniserer ved å sende digitale pulser. Det gjør at det ikke spiller noen rolle hvor sterk pulsen er så lenge den er tilstede, og at det er lett å forsterke/reptere signalet. Det vanligste er å la den informasjonsbærende enheten være frekvensen av pulser, men det er også mange andre muligheter. Cariani deler opp nevromorf koding i tre hovedgrupper (Cariani 1997):

- ▶ Kanal koding (channel) ser på hvilke nevroner (kanaler) som svarer på en gitt stimuli.
- ▶ Temporært mønster koding (temporal pattern) ser på hvordan nevronene sender ut sine pulser. Om det er i korte ”bursts” og eventuelt pausen mellom slike ”bursts”.
- ▶ Tid motatt koding (time-of-arrival) ser på når nevronene sender ut pulser etter en stimuli.

Digitale pulser er likevel svært forskjellig fra digitale signaler i moderne elektronikk. Mens digitale systemer har et eller flere klokkesignaler som styrer når endringer kan skje, kan nevronene sende ut pulser når som helst. Pulsene er derfor analoge i tid og digitale i verdi. Nevromorf koding blir derfor veldig spesiell, siden den er både digital og analog. På et senere tidspunkt merkes adresse-hendelsen med et digitalt tidsstempel. Analog til digital konvertering kan gjøres enkelt ved å telle opp antall pulser i et gitt tidsrom, dersom informasjonsbæreren er pulsfrekvensen.

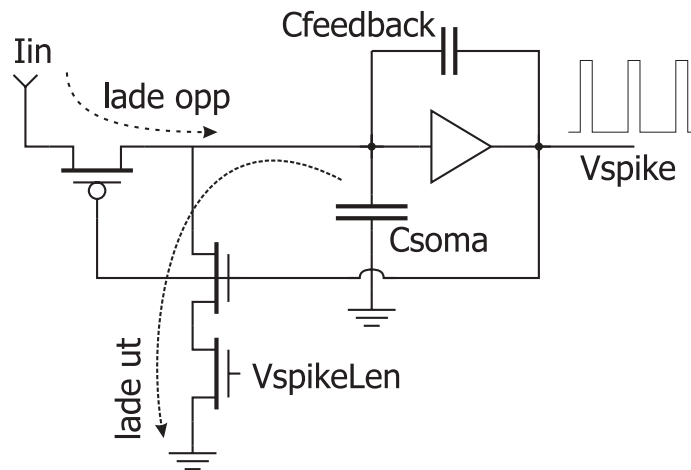
For kameraet ligger den viktige informasjonen i hvilket nevron som først sender ut en puls. Det vil være det nevronet tilkople raden/kolonnen med mest endringer, og derfor størst strøm inn og størst pulsfrekvens ut. I de fleste tilfeller vil dette være det nevronet som er plassert i gravitasjonssentrum til objektet som observeres.

Det er flere typer nevroner å velge mellom. En type er adaptive nevroner, som i likhet med nevronene i biologien har en pulsrate som avtar ved konstant stimuli. Siden fotokretsen allerede er adaptiv, er det ikke nødvendig med slike nevroner. Det finnes også avanserte nevroner som i stor grad modellerer oppførselen til biologiske nevroner, men disse er ikke særlig kompakte. Jeg har valgt å bruke ”integrate-and-fire” nevron kretsen, som er en enkel modell av de biologiske nevronene. Den er også ganske kompakt, men dette er avhengig av hvor store kondensatorer man velger å lage. Fordi nevronene er koplet sammen til en ny type winner-take-all krets, har jeg laget store kondensatorer (forklart senere), og nevronene blir dermed ikke så kompakte. Fordi nevronene kun finnes langs kanten av matrisen, er ikke dette et stort minus.

6.2 Integrate-and-fire nevront

En relativt enkel krets som modellerer biologiske nevronter er *integrate-and-fire* nevrontkretsen introdusert av C. Mead i boken Analog VLSI and Neural Systems (Mead 1989). I likhet med at nevrontene i kroppen lader opp en spenning over celleveggen nâr de får input, lader denne kretsen opp en kondensator. Nevrontene i kroppen har så en selvforsterkende reaksjon ved nådd terskelspenning på omtrent -40 mV. De sender ut en kraftig puls, en "spike". Tilsvarende generer kretsen også en kraftig digital puls når terskelspenningen er nådd. For både det biologiske nevrontet og nevront-kretsen, har input lite å si mens nevrontet sender ut pulsen.

Det er laget mer avanserte nevronter som mer nøyaktig simulerer biologiske nevronter, men for kameraets bruk, er "integrate-and-fire" nevrontet tilstrekkelig. Det er også et poeng at kretsen ikke er for komplisert, også dette lært fra naturens side. Man bruker det man har på best mulig måte. Det er ikke et mål i seg selv å få kameraet til å virke nøyaktig likt som øyet, det er tross alt forskjell på biologisk materiale og silisium. Det er idéene og prinsippene vi ser etter og prøver ut.



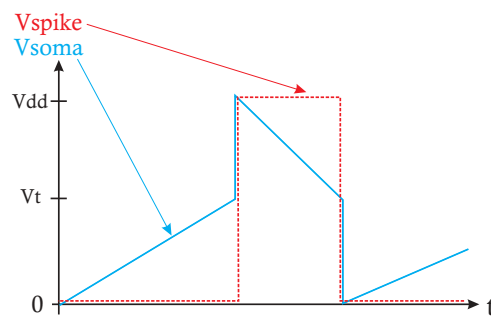
Figur 17 - integrate-and-fire nevrontet

Integrate-and-fire nevront kretsen er vist i Figur 17 og har følgende virkemåte: Inngangen til nevront kretsen er strømmen I_{in} . P-transistoren på inngangen er helt åpen med $V_{gate} = V_{spike} = 0$ V. De to transistorene i serie er lukket. Den øverste har $V_{gate} = 0$ V $\rightarrow V_{gate-source} \approx -0,7$, så den er helt lukket. Den andre har en fast bias-spenning som slipper gjennom en liten strøm, men siden de står i serie går det ingen strøm gjennom denne grenen. Hele I_{in} går derfor til kondensatoren C_{soma} .

Etter hvert som strømmen siger inn på C_{soma} , stiger spenningen til den når terskelspenningen til forsterkeren, $V_T \approx 2,5$ V. Terskelspenningen er avhengig av størrelses forholdet mellom n- og p-transistorene den er bygget opp av. På tilsvarende måte som det biologiske nevrontet, starter da en selvforsterkende reaksjon som er uavhengig av inngangen. Utgangen V_{spike} går da høy ($V_{spike} = 5$ V) og inngangen blir koplet fra ved at pMOS transistoren slås av. I tillegg koples V_{spike} gjennom C_2 slik at

V_{soma} stiger til ca. 5V. Denne delen er den selvforsterkende delen, i og med at V_{soma} nå er langt større enn rett før den nådde terskelspenningen.

$V_{spike} = 5V$ fører også til at nMOS transistoren i lekkasjedelen åpnes slik at V_{soma} gradvis synker igjen. Størrelsen av lekkasjestrømmen bestemmes av $V_{spikeLen}$, som også bestemmer hvor lenge en spike varer. Når V_{soma} har sunket ned til terskelspenningen, går V_{spike} til 0V, og resterende ladning på C_{soma} koples gjennom C_2 til utgangen av forsterkeren, slik at V_{soma} også går til 0V. Neste sykel starter så igjen med at inngangen åpnes og strømmen I_{in} begynner igjen å lade opp C_{soma} .



Figur 18 - hvordan en puls genereres av integrate-and-fire nevronet

6.3 Winner-take-all kretser

En nyttig funksjon for mange kretser med sensorer, er å finne den største av flere innganger, da man vil konsentrere oppmerksomheten om denne. En *winner-take-all* (WTA) krets gjør nettopp dette. Ved å gjøre denne beregningen på brikken i parallell blir den mye raskere enn å beregne serielt i en generell prosessor senere.

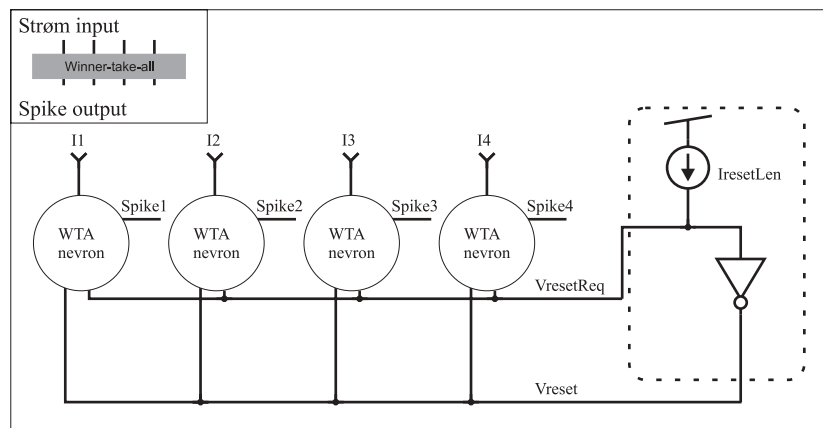
Den klassiske WTA kretsen, introdusert av (Lazzaro et al. 1988) har en utgang som er proporsjonal med logaritmen av den største inngangen. Den skalerer med $O(N)$, og involverer kun to transistorer pr. inngang, pluss en felles for hele kretsen. Resultatet leses vanligvis ut som spenning, men kan enkelt gjøres om til en strøm ved å la denne spenningen gå inn på en transistor.

Som tidligere nevnt, er det ønskelig at utgangen av kretsen er pulser (slik som ganglion cellene sender nervepulser videre inn til hjernen). En enkel løsning på dette ville vært å la hver utgang av den klassiske WTA kretsen gå inn til et *integrate-and-fire* nevron for å få generert pulser.

Tidsdomene WTA kretsen presentert i neste avsnitt kombinerer disse to funksjonene i en og samme krets ved å gjøre enkle modifikasjoner av I&F nevronet. I forhold til to kretser sparer dette plass.

6.4 Tidsdomene WTA av integrate-and-fire nevroner

Ideen om å la nevronet som først sender ut en puls nullstille de andre nevronene ble tenkt ut, og winner-take-all kretsen ble utviklet. Jeg kunne ikke finne noen liknende kretser. I utviklingen av kretsen kunne det virke som den hadde større nøyaktighet enn andre winner-take-all kretser, så lenge strømmene var svært små ($< 1 \text{ nA}$).



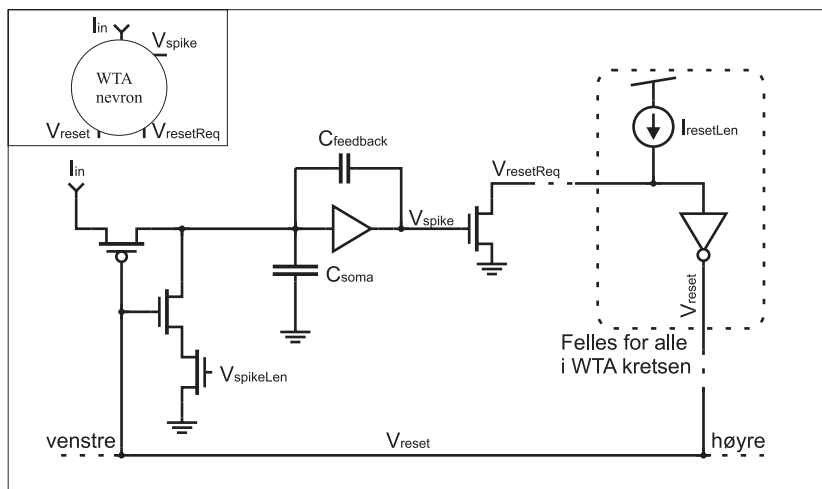
Figur 19 - en winner-take-all krets av integrate-and-fire nevroner.

”Winner-take-all” (WTA) kretsene fra Figur 4 er konstruert som vist i Figur 19 ovenfor, her vist med fire nevroner. Dens virkemåte er enkel: Det første nevronet som sender ut en puls nullstiller ikke bare seg selv, men alle nevronene i kretsen. Så lenge dette nevronet har størst inngangsstrøm, vil det alltid være det første til å sende ut en puls, og dermed også det eneste som får sendt ut sine pulser. Det diskriminerer de andre nevronene. Vi kan med andre ord si at det vinner.

En markant forskjell fra andre ”winner-take-all” kretser er at konkurransen foregår i tidsdomenet. Det vil si at det er den hurtigste (i tid) som vinner, fremfor den største strømmen/spenningen. Når det likevel er nevronet med den største inngangsstrømmen som vinner, så foregår selve konkurransen i tid, og det er derfor jeg har kalt den en ny type WTA krets.

Winner-take-all kretsen har også en maks funksjon, ved at den ikke bare velger ut den største, men også sender ut pulser med en frekvens som er proporsjonal med inngangsstrømmen.

6.4.1 Et nevron i winner-take-all kretsen



Figur 20 - implementasjon av nevro-WTA kretsen.

Hvert av nevronene i WTA kretsen er implementert som vist i Figur 20. Dette er en modifikasjon av integrate-and-fire nevronet beskrevet tidligere. Kretsen virker som følger:

Inngangsstrømmen inn på I_{in} lader gradvis opp kondensatoren C_{soma} . Utgangen V_{spike} er lav. Når spenningen over C_{soma} når terskelspenningen til forsterkeren (to invertere i serie) vil utgangen V_{spike} brått skiftes til V_{dd} , og vi har en puls. Skiftet av spenningen fra jord til V_{dd} koples gjennom tilbakekopplingskondensatoren $C_{feedback}$, og gjør at spenningen over C_{soma} gjør et hopp fra den akkurat nådde terskelspenningen til et sted nær V_{dd} . Vi har nå en stabil tilstand der C_{soma} har en spenning langt over forsterkerens terskelspenning.

V_{spike} høy fører til at den globale $V_{resetReq}$ trekkes lav. Denne linjen holdes normalt høy av strømkilden $I_{resetLen}$, og former sammen med de andre nevronene en såkalt ”wired” NOR. Det vil si at et hvert nevron som sender ut en puls vil trekke denne linjen lav. Inverteren gjør at V_{reset} går høy, som fører til at inngangen I_{in} til nevrontet koples fra, og C_{soma} lades gradvis ut, begrenset av $V_{spikeLen}$. Siden V_{reset} er felles for alle nevronene betyr det at alle nevronene lades ut.

Når vinner nevronet, det som sender ut en puls, har fått sin spenning over C_{soma} redusert til terskelspenningen for sin forsterker, vil V_{spike} skifte til jord, som vil si at pulsen er ferdig. Dette fører også til at $V_{resetReq}$ kan begynne å lades opp igjen av strømkilden $I_{resetLen}$. Dette vil etter en stund gjøre at V_{reset} skifter til lav, og nevronene kan begynne å lade seg opp mot neste puls. Med andre ord er en ny syklus startet.

En annen effekt av at V_{spike} går lav er at den ekstra ladningen som ble ”lånt” gjennom C_2 leveres tilbake. Spenningen C_{soma} for vinner nevronet skifter derfor hurtig ned fra forsterkerens terskelspenning til jord. Dette gjør at vinner nevronet ikke bruker lenger tid enn resten av nevronene på å lades ut.

6.4.2 Oppløsning og skalering

I areal skalerer denne kretsen som $O(N)$. Antall nevroner (deltakere) i kretsen begrenses likevel på grunn av de globale linjene. Hvis disse blir veldig store, og får mange nevroner koplet til seg, kan det bli mye kapasitans fra linjen selv og fra tilkoblede transistorers "drain" kapasitans, og dette gjør kretsen tregere. En modell av nullstillingsdelen av winner-take-all kretsen er vist under.



Figur 21 - den globale nullstillingsdelen består av en "wired NOR" og en sterk inverter.

Et hvert nevron kan starte utladningsprosessen ved å trekke den globale linjen $V_{resetReq}$ lav. Det er kun når et nevron sender ut en puls (eng: "spike") at dette skjer. Alle nevronene er koplet til $V_{resetReq}$ linjen i en kopling som kalles "wired NOR". Det vil si at en felles strømkilde trekker linjen opp, og at alle nevronene kan trekke linjen ned, da de er sterkere enn strømkilden. $V_{resetReq}$ er så koplet til en kraftig inverter, som trekker V_{reset} linjen høy så raskt som mulig slik at utladingen begynner. Når vi kopler flere nevroner til kretsen, vil den ha større "fan-in" og "fan-out", som det heter når man snakker om logiske porter. Inverteren er laget særlig kraftig for å kompensere for stor fan-out (52 tilkoblede nevroner i den større av WTA kretsene). For å forhindre at $V_{resetReq}$ tar lang tid å trekke lav er n-transistorene som trekker den ned laget store, og strømkilden som trekker den høy justeres slik at den ikke er alt for sterk.

Oppløsningen til et winner-take-all nettverk er hvor små forskjeller det klarer å diskriminere mellom. For denne kretsen vil det være hvor like inngangsstrømmene må være for å få to vinnere. For at kun et nevron skal vinne, må betingelsen i formelen under oppfylles:

$$T_1 - T_2 > t_d \quad (13)$$

T_1 er tiden vinneren starter sin puls, og T_2 er tiden neste nevron ville startet sin puls. Forsinkelsen i kretsen, t_d , er tiden fra vinner nevronet starter sin puls til alle andre nevroner er forhindret fra å sende ut pulser, som vil være når utladningsprosessen har begynt. Som viser, er forsinkelsen i (13) kretsen avgjørende for dens evne til å skille ut en klar vinner.

Oppløsningen til winner-take-all kretsen er avhengig av hvilket nivå strømmene har. Større strømmer fører til at nevronene lades opp raskere, slik at $T_1 - T_2$ blir mindre for to store strømmer med samme relative forskjell som to mindre strømmer. Dette er vist i formelen nedenfor, som viser tiden det tar før et nevron sender ut en puls:

$$T = \frac{V_T C}{I} \quad (14)$$

Her er V_T terskelspenningen til forsterkeren, C den kombinerte kapasitansen av C_{soma} og $C_{feedback}$ og I strømmen inn på nevronet. Ved å kombinere (13) og (14) får vi et uttrykk for hvor små strømmer kretsen kan skille mellom:

$$\left(\frac{1}{I_2} - \frac{1}{I_1} \right) > \frac{t_d}{V_T C} \quad (15)$$

Forskjellen uttrykt som prosent $p = 1 - I_2/I_1$ av inngangsstrømmen til vinner nevronet, får vi:

$$p > \frac{k}{1+k}, k = I_1 \frac{t_d}{V_T C} \quad (16)$$

Som vi skal se senere, er verdien av $C = 1,56$ pF og $V_T \approx 1,9$ V i implementasjonen av kretsen. Målinger antyder at $t_d \approx 1$ ns. Dette gir at kretsen, for strømmer i størrelsesorden 10 nA, kan skille mellom to strømmer bare 0,0007% forskjellige. Siden dette er området kretsen brukes i, er den meget godt egnet.

6.4.3 Nullstilling av nevronene

Nevronene som winner-take-all kretsen består av, lades ikke ut med lik hastighet. For det første vil det være forskjeller i utladningsmekanismen på grunn av størrelsesforskjeller mellom transistorer i moderne prosesser. Disse kan være store, så minimumstransistorer bør unngås.

For det andre vil nevronet som vinner ha en større V_{soma} spenning enn de andre nevronene, faktisk dobbelt så stor. Dette gjør at utladningstransistorene (markert i Figur 17) vil ha mye høyere *drain-source* spenning enn resten. Early effekten vil da gjøre at utladningsstrømmen er større for vinner nevronet. Dette var synlig på simuleringer, og påvirket konstruksjonen av kretsen. En måling som tydelig viser at utladningsraten er større når et nevron vinner enn når det taper, er vist i Plott 9 på side 62. Selv om vinner-nevronet har en større ladning å lade ut, vil den lånte ladningen fra $C_{feedback}$ kondensatoren leveres tilbake samme vei. Det er dermed kun en del av ladningen som lades ut gjennom de to transistorene i serie. Som både simuleringer og målinger viser, lades vinner nevronet ut på kortere tid enn resten av nevronene.

6.4.4 Nødvendigheten av todelt nullstilling

En interessant begrensning med tids-domenet WTA er nødvendigheten av å dele opp reset-linjen i to deler; $V_{resetReq}$ og V_{reset} . Dette gjør at tiden nevronene nullstilles, når et nevron har vunnet, kan uttrykkes ved:

$$t_{total\ nullstillings\ tid} = t_{vinner\ pulslengde} + t_{reset}$$

Tiden t_{reset} , som styres av $I_{resetLen}$ (se Figur 19), gjør at nullstillingstiden ikke er styrt av et enkelt nevrons pulslengde. Dersom denne settes lik null, vil vi kunne få ”tjuvstart effekten”. Dette er en effekt vi får dersom vinner nevrons pulslengde skulle styre nullstillingen av alle nevronene, slik den enkleste implementasjonen av kretsen ville ha sett ut, med alle nevronene direkte koplet til reset-linjen. Effekten gjør at ikke alle nevronene blir fullstendig utladet, fordi utladningstiden er for kort, igjen fordi vinnerens pulslengde er kortere enn resten. Det er da mulig at et nevron med nesten like stor inngang som den rette vinneren, bygger seg opp en ”tjuvstart ladning” for hver nullstillingsprosess. Den vil eventuelt kunne vinne, som vist i måleeksempler.

En ting som karakteriserer ”tjuvstart effekten” er at nevronet eller nevronene som jukser vil vinne nøyaktig en gang hver n -te gang. Dette er på grunn av at når et nevron som egentlig ikke skulle vunnet til slutt vinner, så blir det fullstendig utladet, fordi det da er det selv som bestemmer utladningstiden og fordi det lades raskere ut. Det vinner derfor ikke to ganger på rad.

Ufullstendig utladning oppstår som følge av to ting:

- ▶ Vinner nevronet lades raskere ut enn de andre nevronene
- ▶ Forskjellig transistorstørrelser gjør at noen nevroner lades raskere opp/ut enn andre, som vises tydelig på målinger.

Ved å øke t_{reset} får vi resultatet vist i Plott 10, som er normal oppførsel. Sammenliknet med Plott 9 ser vi at nevronene nå har en hvileperiode etter hver puls, en periode der nevronene ikke kan begynne å lades opp. Dette gjør at vi er sikre på at alle nevronene er fullstendig nullstilt når en ny ”konkurranse” begynner.

$I_{resetLen}$ må stilles inn i forhold til utladningsraten (styrt av $V_{spikeLen}$). Lang pulslengde gjør at forskjellen mellom nevronet som først lades ut og det som sist lades ut blir stor. Følgelig må treset være lang. Kort pulslengde gjør at alle nevronene lades ut raskere, og dermed at forskjellen mellom nevronet som først og sist nullstilles blir kortere, slik at også t_{reset} kan gjøres kortere.

6.4.5 Inaktivitets (timeout) nevronet

Dersom endringspikslene hadde brukt en antikorrelasjonskrets, slik at de *leverte* strøm ved endringer og ikke ellers, ville det i teorien ikke gått noe strøm inn til nevronene ved statisk bilde. På grunn av at endringspikslene trekker strøm ved statisk bilde, og ikke ellers, er de koplet opp som beskrevet i avsnitt 5.8.1. Som Figur 16b (side 46) viser, er endringspikslene og nevronene effektivt parallellkoplet, og det vil derfor alltid gå en viss strøm inn til winner-take-all kretsen.

Naturlig nok vil nevronene av denne grunn også generere pulser, noe som ikke er ønskelig siden kameraet er en endringsdetektor, og ingenting endrer seg. Dette vil kunne kategoriseres som en falsk alarm.

Inaktivitetsnevronet er et ekstra nevron koplet til winner-take-all kretsen. Det skiller seg ut ved at det får sin inngang fra en transistor som fungerer som en strømkilde, og ikke fra en rad/kolonne. Siden radene og kolonnene har hvert sitt WTA, har de også hvert sitt inaktivitetsnevron.

Virkemåten til inaktivitetsnevronet er enkel: Inngangsstrømmen I , som er styrt utenfra brikken, gjør at nevronet vil sende ut en puls etter tiden T (se formel (14)) hvis det får sjansen (ingen nevroner har vunnet før tiden T). Som for alle nevronene, vil denne pulsen gjøre at alle nevronene blir nullstilt. Inaktivitetsnevronet gjør derfor at hvis alle nevronene har fått små inngangsstrømmer, så vil de aldri få sende ut pulser. Så lenge inaktivitetsnevronet er det eneste som sender ut pulser, vet vi at ingen endringer har funnet sted.

Inaktivitetsnevronet kan også benyttes til å filtrere bort små endringer, som mest er å betrakte som støy. Dette gjøres enkelt ved å tilføre mer strøm, som er ekvivalent med å korte inn tiden før nevronene bli nullstilt (eng. timeout). Når inaktivitetsnevronet får mer inngangsstrøm, må de andre nevronene ha en enda større inngangsstrøm for å nå sin terskelspenning før de blir nullstilt. Med andre ord må flere av endringspixelene slutte å trekke strøm, slik at det blir mer overskuddsstrøm til nevronene. Dette kan bare skje dersom flere endringspixeler samtidig registrerer at en endring har funnet sted. Dette vil vanligvis bety at et større objekt beveger seg i synsfeltet til kameraet. Små objekter vil bare aktivere få endringspixeler, og overskuddsstrømmen inn til nevronene blir da ikke stor nok til at de kan vinne over inaktivitetsnevronet.

Siden inaktivitetsnevronet sender ut pulser i lik linje med de andre nevronene, er det veldig enkelt å holde styr på når kameraet er i "hvilestilling". Så lenge det ikke skjer noe av interesse, vil inaktivitetsnevronet jevnlig sende ut pulser. Vi vet da at kameraet er aktivt og at sannsynligvis fungerer. Vi vet også at det ikke skjer noe av nevneverdig interesse, og eventuelt tilkoplet utstyr kan settes i "stand by" modus eller slås helt av. For en person som skal følge med på flere skjermer samtidig, der det den meste av tiden ikke skjer noe, vil dette være en stor fordel. Bare når det er noe som rører på seg vil TV-skjermene vise bilde og tilkalle vaktens oppmerksomhet. Man kan også kople til en alarm eller lignende for å være sikker på at vekten ser på bildet.

6.5 Digital støy fra nevronene

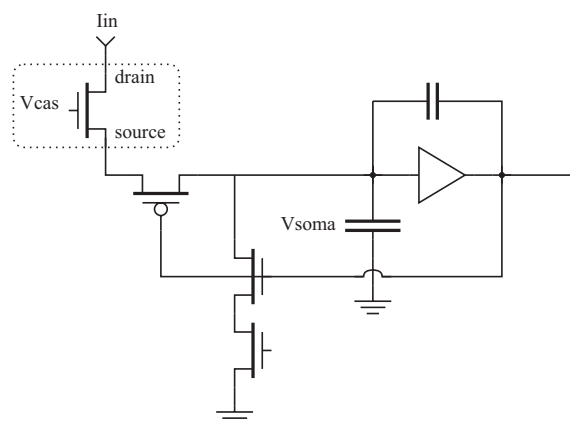
Det vil uunngåelig genereres en del digital støy når nevronene sender ut pulser. Denne støyen kan forplante seg bakover i systemet til de følsomme analoge fotokretsene. I kameraet gjøres det to ting for å begrense effekten av denne støyen.

Det første som er gjort er å øke størrelsen av strømmene i radene og kolonnene, og dermed også i endringspixelene. Svitsjestøyen fra nevronene har samme størrelse uavhengig av inngangsstrømmens størrelse, som kun gjør at nevronene blir raskere ladet opp. Større strøm i radene og kolonnene bedrer derfor forholdet mellom signal og støy. På grunn av egenskapene til winner-take-all kretsen, bør likevel ikke strømmen være alt for stor, da oppløsningen vil bli dårligere. Som formel (14) viser, kan dette kompenseres for ved å lage store kondensatorer, og dette er gjort i kameraet.

Det andre som er gjort er å minske absoluttverdien av støyen. På inngangen til nevronet er det plassert en transistor med en passende forspenning, som vist i Figur 22. En transistor plassert slik kalles en regulert kaskode(transistor), og har vanligvis som oppgave å redusere *drain-source* spenningsvariasjoner i strømspeil (Serrano et al. 1994). Kaskodetransistoren i denne kretsen fungerer som en *source* følger. Det vil si

at så lenge gate-spenningen holdes fast, vil spenningen på *drain* terminalen holdes fast mens source terminalen styres av strømmen.

Formlene for strøm gjennom en transistor (1) og (2) viser oss at dersom *gate*-spenningen er fast så er strømmen gjennom en transistor styrt av *drain-source* spenningen (V_D og V_S i formelen). Omvendt kan man la *drain-source* spenningen styres av strømmen I_{ds} . Dersom *drain* allerede ligger fast (koplet mot strømkilden for raden/kolonnen), er det bare *source* terminalen som styres av strømmen.



Figur 22 - en regulert kaskode transistor (åpen transistor med fast gate-spenning) er lagt inn for å redusere digital støy fra nevronene bakover i systemet

Svitsje støyen begrenses på følgende måte. V_{cas} settes lavt, men ikke så lavt at V_{soma} blir forhindret fra å nå terskelspenningen, for eksempel 3V. V_{soma} lades opp og vil etter hvert nå nevronets terskelspenning og gå rett til Vdd. Source og drain blir da byttet om for kaskodetransistoren siden en n-transistor har source terminalen ved laveste spenning. Den nye source terminalen har nå en spenning som ligger høyt, over 4V. Siden den nye source er høyere enn gate-spenningen, blir gate-source spenningen på transistoren negativ, og transistoren er helt lukket. Ingen forstyrrelser skal da forplante seg tilbake fra nevronet gjennom transistoren.

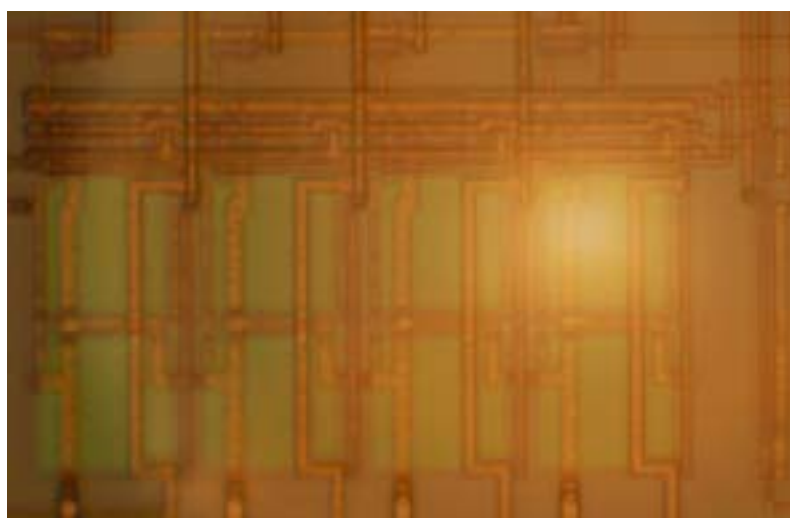
Det kan likevel forekomme noen forstyrrelser som kopler seg kapasitivt fra source til gate og videre fra gate til drain, men disse er små og kan godt ignoreres.

6.6 Utlegg

Utlegg og fotografi av winner-take-all testkretsen med fire nevroner er vist. Mest markant er de to store kondensatorene C_{soma} og $C_{feedback}$ for hvert nevron. Øvrige kretselementer er små i sammenlikning med disse. Ytre mål for kondensatorene er $33,0 \mu\text{m} \times 34,7 \mu\text{m}$, så hver av dem er nesten like store som en endringspiksel.



Figur 23 - Utlegg av WTA



Figur 24 - Fotografi av WTA. De store grønne feltene er kondensatorene C_{soma} og $C_{feedback}$.

6.7 Målinger av winner-take-all kretsen

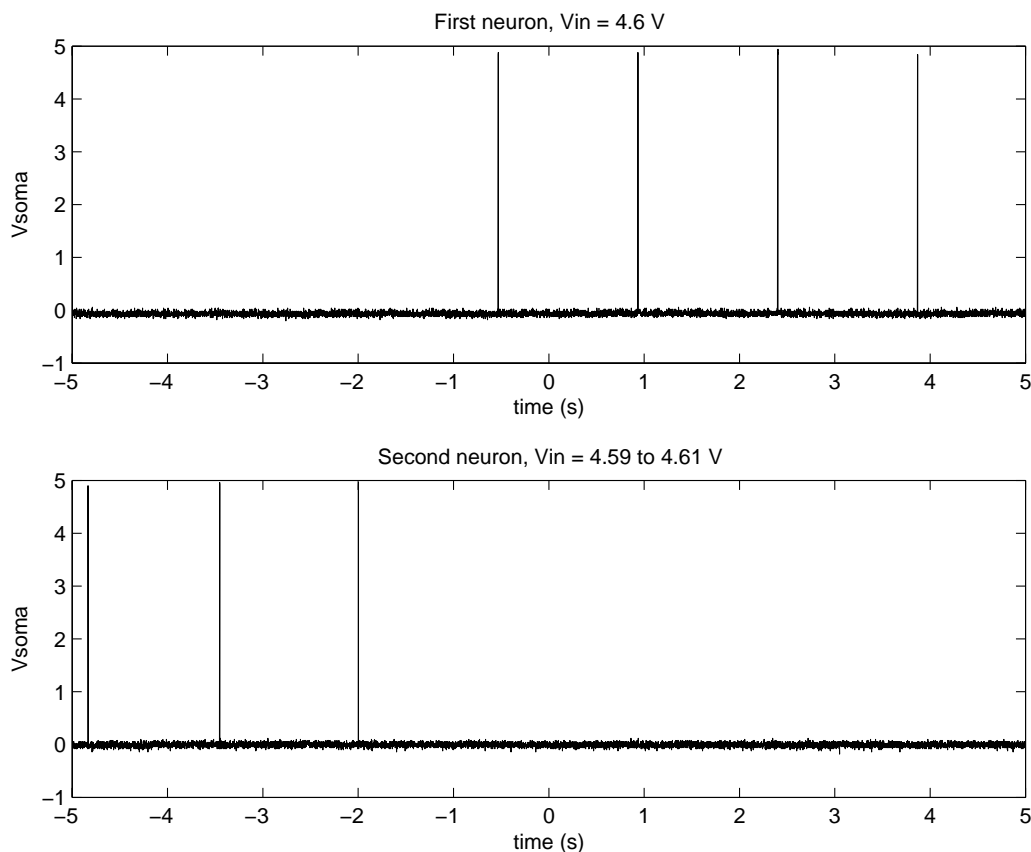
På brikken finnes det to store WTA kretser, en 43 nevroners krets og en 52 nevroners krets. Årsaken til at kretsene har flere nevroner enn det er rader og kolonner (kameraet har 42 rader og 51 kolonner) er at hver krets har et inaktivitetsnevron i tillegg. Dette er forklart tidligere.

Av hensyn til testing er det også lagt inn en fire nevroners utgave av winner-take-all kretsen på brikken, som vist i Figur 19. Denne testkretsen kontrollerte strømkilder (p-transistorer i metning) som inngang til hvert nevron. Det er laget målepunkter for

spenningen C_{soma} på alle fire nevronene. Dette gjør det enkelt å sammenlikne nevroner med oscilloskopet. Man kan også se på pulsene ut fra nevronene, men dette gir oss mindre informasjon om hva som skjer. De globale linjene $V_{resetReq}$ og V_{reset} har også målepunkter (også i hovedkretsen).

6.7.1 Glidende overgang

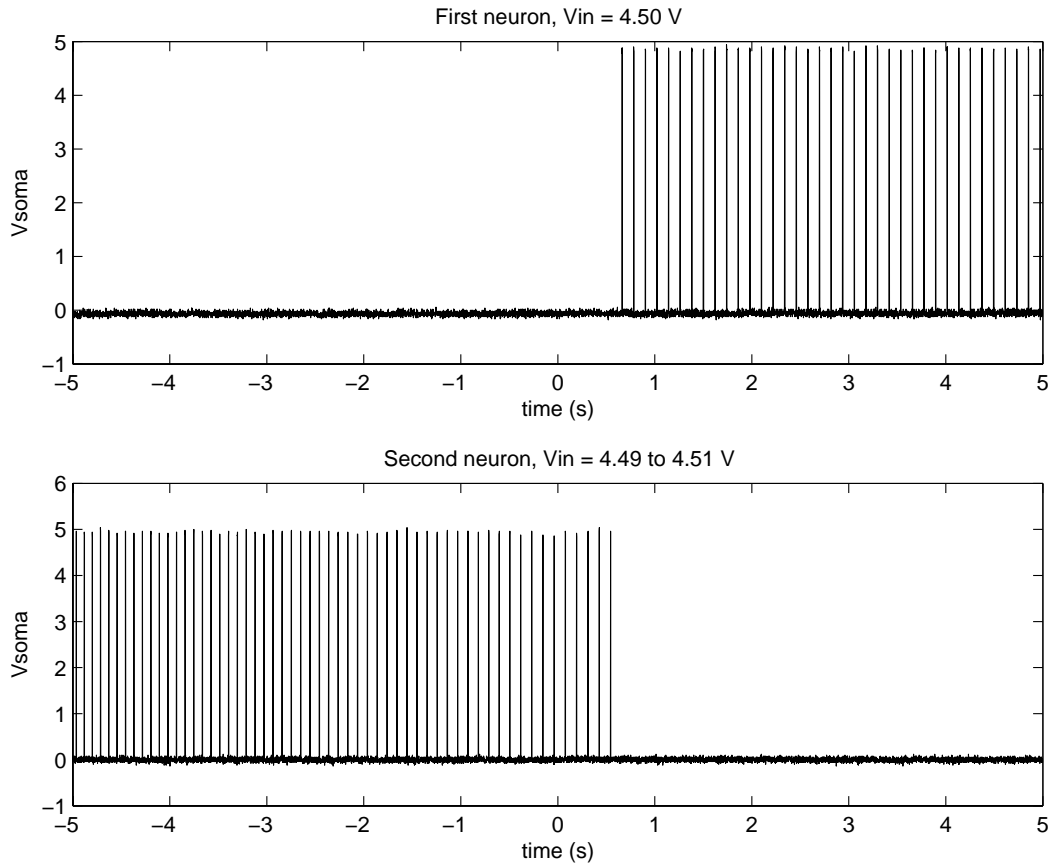
På laboratoriet utførte jeg to ”glidende overgang” tester. Det vil si at et nevrns inngangsstrøm står fast, mens strømmen til det andre varieres fra å ha mer strøm til å få mindre strøm. Nevronet som har fast strøm inn vil da ta over som vinner etter hvert som nevronet som stadig får mindre strøm taper. Et av formålene med en glidende overgang er å se om vi på noe punkt kan få to vinnere, som selvfølgelig er uønsket. Analysen av nevro-WTA viser at oppløsningen blir dårligere for større strømmer. Det er laget glidende overgang plott for to størrelser strømmer. Resultatet er vist i Plott 7 og Plott 8.



Plott 7 - glidende overgang 1. Nedre nevron får gradvis mindre strøm, og taper etter hvert for øvre nevron. Plottet viser pulser i et fire-nevrns WTA. Beregnet strøm for øvre nevron er ca. 2,8 pA.

I plottet ser vi at det ikke finnes noe overlapp, altså en periode der begge nevroner vinner. Neste plott (Plott 8) har større strøm, ca. 10 ganger mer, og dermed også

dårligere oppløsning. Siden strømmen er så liten som ≈ 27 pA, er oppløsningen fremdeles meget god, og heller ikke her får vi noe overlapp.

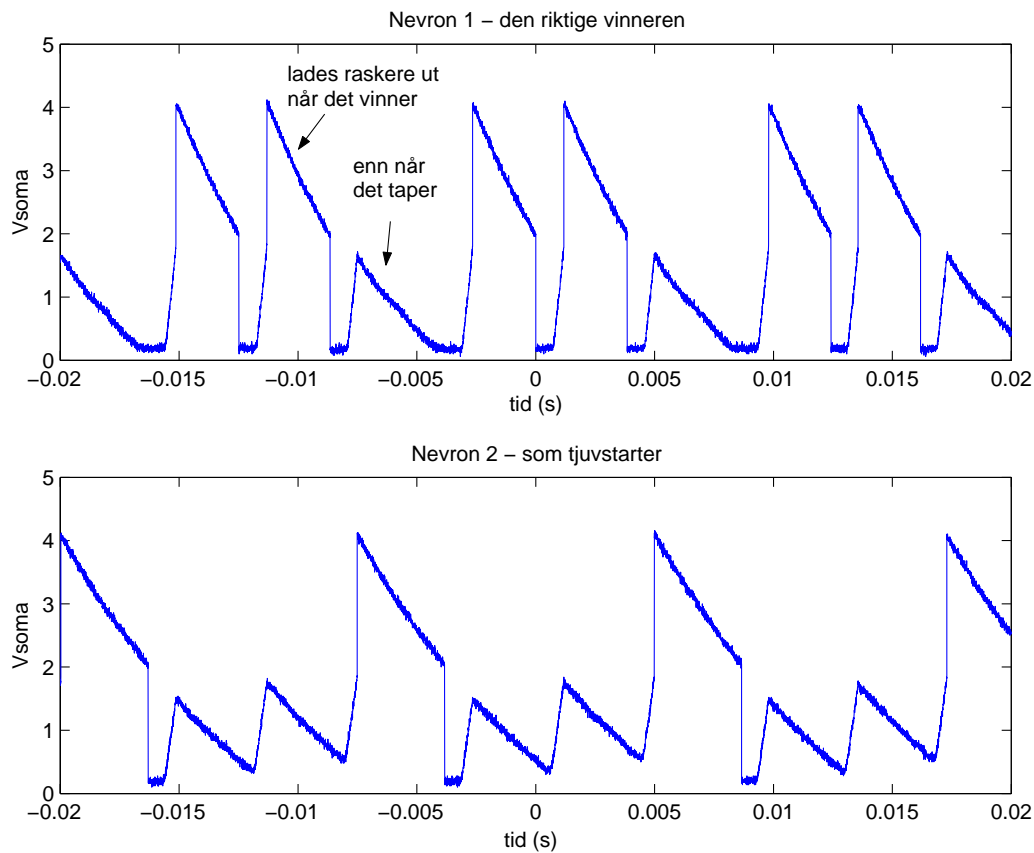


Plott 8 - glidende overgang 2. Nedre nevron får gradvis mindre strøm (periode på 3 s) mens øvre nevron holdes stabilt på ca. 27 pA. Når nedre nevron får mindre strøm enn øvre nevron, vinner øvre nevron.

Dette plottet viser også kretsens maks-funksjon, som kan se som en synkende frekvens av pulser i det nedre plottet, der strømmen minkes.

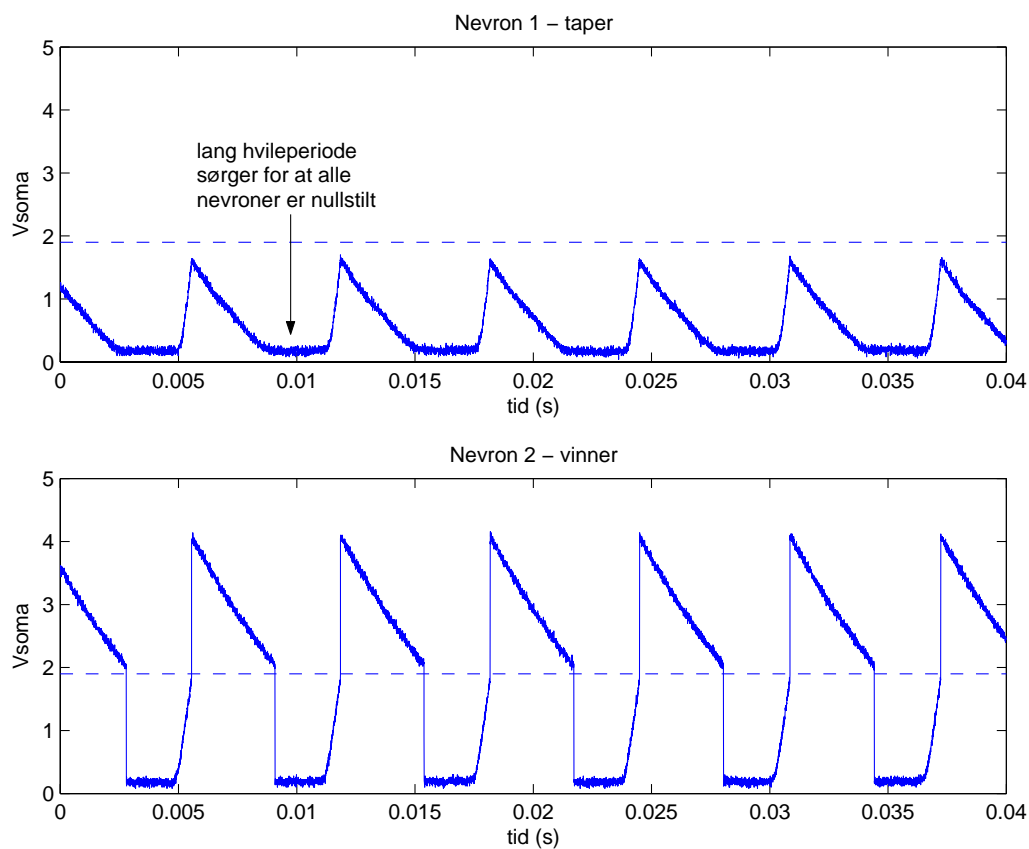
6.7.2 Eksempler på ufullstendig nullstilling

Et eksempel på "tjuvstart effekten" er vist i Plott 9. To nevroner har nesten lik oppladningstid T fra formel (14). Her er $t_{reset} \approx 0$ s, slik at nullstillingstiden bestemmes av det vinnende nevrons puls lengde alene.



Plott 9 - et eksempel på "tjuvstart effekten". Nedre nevron blir ikke helt nullstilt ved hver puls, og opparbeider seg litt mer ladning for hver konkurranse. Dette gir nedre nevron en tjuvstart som fører til at det vinner selv om det ikke burde vunnet. I dette plottet vinner nedre nevron hver tredje konkurranse.

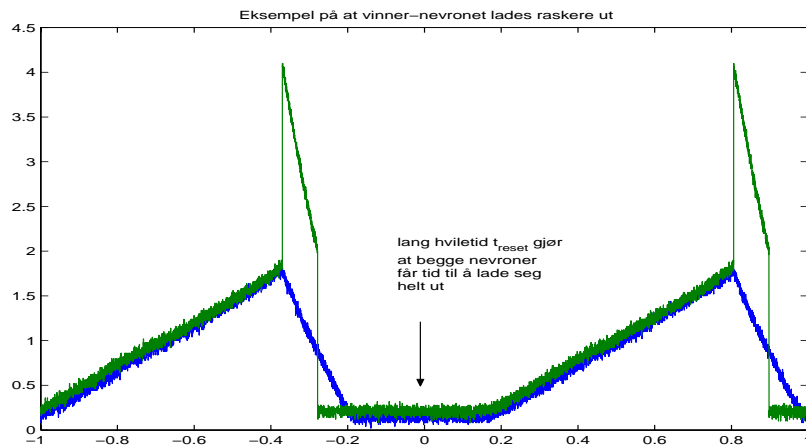
Ved å øke t_{reset} får vi resultatet vist i Plott 10, som er normal oppførsel. Sammenliknet med Plott 9 ser vi at nevronene nå har en hvileperiode etter hver puls.



Plott 10 - normal operasjon for winner-take-all kretsen

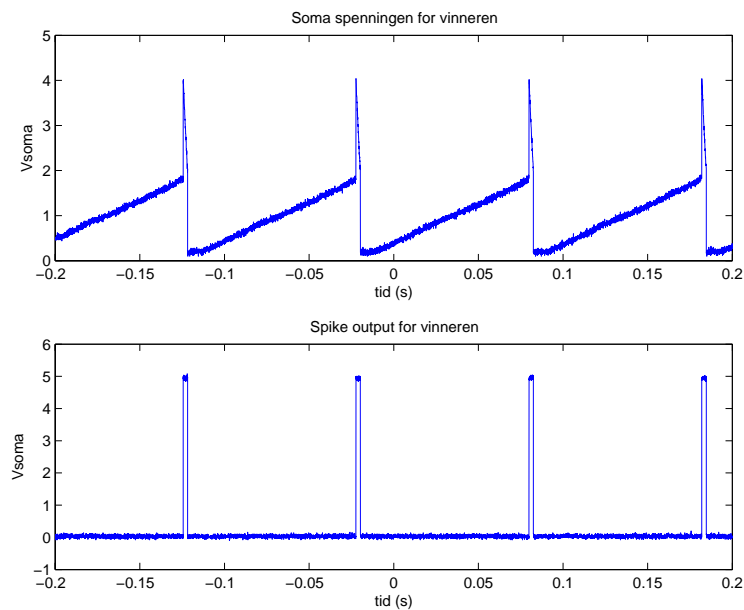
Litt overraskende vinner nevron nr. 2 (som var taperen tidligere), selv om strømmene skulle være uendret i forhold til Plott 9. Dette kan komme av at noen nevroner begynner å lade opp litt tidligere enn andre, sannsynligvis pga. forskjeller i transistorstørrelser. Når hvileperioden er over åpnes nevronene for strøm inn. Selv om dette skjer meget raskt, vil noen nevroner åpnes litt før andre pga. forskjeller i transistorstørrelser.

Plott 11 viser enda tydeligere hvorfor lang nok t_{reset} (hviletid) er viktig for korrekt operasjon av winner-take-all kretsen.



Plott 11 - viktigheten av lang nok hviletid

Et eksempel på operasjon ved små strømmer i forhold til pulslengden er vist i Plott 12. Informasjonen i dette plottet kan brukes til å regne ut inngangsstrømmen.



Plott 12 - vinnerens Vsoma og puls output

Strømmen inn til vinner-nevrontet kan regnes ut på to måter.

- Ved å regne ut stigningstallet til V_{soma} og multiplisere med kapasitansen C . På figuren kan det se ut som V_{soma} ikke blir helt 0 V, men dette tror jeg kan skyldes spenningsfølgeren som brukes til å forhindre at avlesning av spenningen forstyrrer. Utregningen for vinneren i Plott 10 er vist under:

$$\frac{dV}{dt}C = I \Leftrightarrow \frac{1,9V}{0,09s} \cdot 1,56pF \approx 33pA$$

(En stigetid på 0,09s pr. puls tilsvarer en frekvens på 11,1 Hz.)

- Ved å se på frekvensen av pulser. Denne er enklere å observere, men pulslengden og t_{reset} bør trekkes fra, siden disse ikke er tatt høyde for i formelen under. Den observerte frekvensen (fra det nedre plottet) er litt lavere (10,1 Hz) og resulterer i at vi beregner en strøm som er litt mindre. V_{Θ} er terskelspanningen

$$f = \frac{I}{V_{\Theta}C} \Leftrightarrow I = fV_{\Theta}C = 10,1Hz \cdot 1,9V \cdot 1,56pF \approx 30pA$$

6.7.3 Måling av forsinkelsen

For å karakterisere kretsen, er det nyttig å vite hvor stor forsinkelsen t_d i formel (13) er. Dette påvirker oppløsningen til kretsen. Dessverre er t_d vanskelig å måle pga. forsinkelser i spenningsfølgere på utgangene og i pad-ene. Den målte forsinkelsen er $V_{resetReq} - V_{reset}$. Ved å øke spenningsfølgerenes bias-strøm, slik at de blir raskere, blir den målte forsinkelsen mindre. Denne strømmen kan ikke økes for mye, da det viser seg å ha en uheldig effekt på kretsene (en krets sluttet å virke). Forsinkelsen på den ene av de to store (53 nevroner) WTA-kretsene er målt til å være

$$t_{forsinkelse} < 4 ns^*$$

(*) Forsinkelsen skal måles fra vinner-nevront sender ut sin puls, til V_{reset} er høy. Dette går ikke, da pulsene fra nevronene går rett ut av pad-en, mens V_{reset} går gjennom to spenningsfølgere. Den målte forsinkelsen er derfor fra $V_{resetReq}$ går lav til V_{reset} er høy, som gir en indikasjon på hva forsinkelsen kan være.

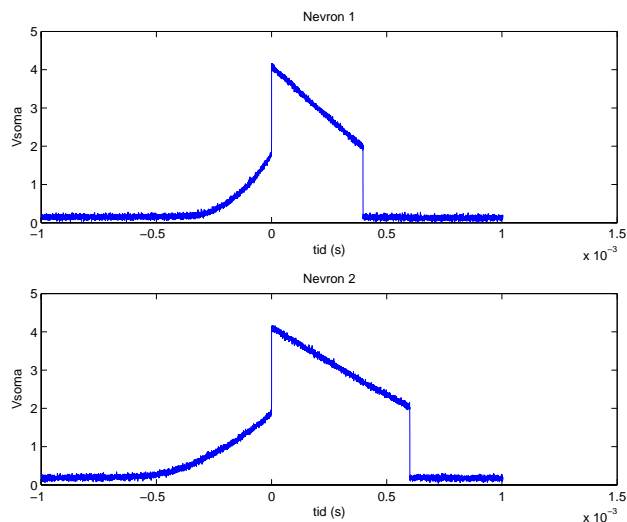
6.7.4 Forsøk på å få flere vinnere ved å øke strømmen

Følgende forsøk er gjort for å se om det er mulig i praksis å få to vinnere. Dette gjøres ved å øke inngangsstrømmene, samtidig som man på oscilloskopet prøver å justere slik at to nevroner når terskelspanningen tilnærmet samtidig. I forsøket endte inngangsspenningene (som styrer p-transistorer som fungerer som strømkilder) på 4,18 V og 4,26 V.

De ulike spenningene, i tillegg til det man kan se av kurvene, viser at det er stor forskjell på nevronene. Den ene begynner bl.a. å lade seg opp tidligere enn den andre. Den andre lader seg opp og ut mye raskere.

Det var ikke mulig i praksis å få to nevroner til å vinne samtidig utenom tilfeldig, forårsaket av støy. I simuleringer var det mulig å få stabilt to vinnere ved 0,08 % forskjell for en strøm rundt 124 nA, mens for en lavere strøm på ca. 8,4 nA diskriminerte kretsen mellom to innganger på bare 0,003% forskjell. Dette stemmer overens med formlene.

I praktiske tester fikk jeg stilt to nevroner så likt at støyen alene bestemte hvilket av de to nevronene som vant. Testene viste da gjerne 10-20 pulser fra et nevron før det skiftet til et annet nevron. Ved noen tilfeller kunne støy føre til at to nevroner vant samtidig, men dette skjedde kun unntaksvis. Det var derfor ikke mulig å måle hvor små strømmen kretsen kunne diskriminere mellom. Et tilfelle der to nevroner vant samtidig er vist i Plott 13.



Plott 13 - unntaksvis kunne to nevroner vinne samtidig som følge av støy.

Plott 13 viser også hvordan størrelsesforskjeller mellom transistorer kan føre til at et nevron begynner å lades opp tidligere enn det andre, samt hvor forskjellig utladningstiden kan være. Det er lite sannsynlig at dette skyldes forskjeller i kondensatorer, som vanligvis har god nøyaktighet.

Beregnet strøm for nevron 2 i Plott 13:

$$\frac{dV}{dt}C = I \Leftrightarrow \frac{1,9V}{0,5ms} \cdot 1,56pF \approx 5,9nA$$

Det viste seg vanskelig å få til større strømmen enn dette, da dette ga maksimal puls rate for kretsen.

7 AER PROTOKOLLEN OG MÅLINGER AV HELE SYSTEMET

7.1 Utgangssignaler fra systemet

En stor forskjell på "kretser" i biologien og mikroelektronikk, er at man i biologien har veldig mange flere forbindelser. Størsteparten av plassen i hjernen vår er forbindelser mellom nevronene og kalles ofte nervebaner. Mange forbindelser går ut av og inn til hjernen, og kan være svært lange, opptil flere meter. Signaler skal kunne gå helt fra tærne og opp til hjernen. Underveis forsterkes og repeteres signalene flere ganger i det som kalles Ranviér nodene.

I mikroelektronikk er det ikke praktisk mulig å lage like mange forbindelser som vi har i biologien. Dette skyldes først og fremst at kretser i silisium er todimensjonale. Riktignok har man noen lag oppå hverandre slik at man kan krysse ledninger, men noen full tredimensjonal struktur har man ikke, så det er ikke mulig å lage et antall forbindelser som er sammenliknbart med det som finnes i biologien.

I hjernen har hvert nevron separate "punkt-til-punkt" forbindelser til tusenvis av andre nevroner. På disse forbindelsene overføres informasjon i form av digitale pulser. Når dette skal lages i mikroelektronikk må vi dra nytte av flere fenomener. For det første er forbindelsene mellom nevronene i store deler av tiden ubrukte. For det andre er ledningene vi har tilgjengelige i mikroelektronikk mye bedre enn de som finnes i biologien. Vi kan både overføre mye raskere og over lengre avstander uten å forsterke opp (repetere) signalene. Kombinasjonen av bedre ledninger og mye dødtid mellom hver puls gjør at nevromorfe signaler er en god kandidat til å tidsmultiplekseres på en felles buss.

Multipleksing er kjent fra datakommunikasjon der flere sendere og mottakere deler et felles medium. Det er flere måter å multiplekse på, der tidsmultipleksing er en teknikk som går ut på å la pulser fra alle nevronene sendes ut på samme sett med ledninger, men til forskjellig tid. Siden det er så mye dødtid mellom pulsene vil det være lett å få til. En utfordring blir likevel hvordan man på mottakersiden skal kunne skille mellom pulser fra ulike nevroner.

I mikroelektronikk er signalene ofte mer sammensatte enn i biologien, for eksempel ved forskjellige kodesystemer. Det mest kjente er binær koding, som ikke trenger noen nærmere forklaring.

7.2 Adresse-hendelse (AER) protokollen

Adresse-hendelse protokollen (Address-Event Representation, (AER)) utnytter dette på følgende måte: Hvert nevron får tildelt en unik binær adresse. Når det ønsker å sende ut en puls (*hendelse*), plasserer det sin *adresse* på en felles buss. På mottakersiden er det da enkelt å se hvilket nevron som sendte ut pulsen, siden det bare er å dekode adressen. Adresse-hendelse protokollen gir oss punkt-til-punkt forbindelser mellom nevroner på en delt buss (Bohaen 2000).

Ved å la utgangene følge adresse-hendelse protokollen spares antall forbindelser i stor grad. Mer bestemt vil man ha et antall ledninger i bussen gitt ved

$$\text{Antall ledninger i bussen} = \log_2 N \text{ (rundet opp)} \quad (17)$$

der N er antall ulike adresser som skal kodes (= antall nevroner). I biologien det være et antall "ledninger" gitt ved $N \cdot \text{antall tilkoplinger pr. nevron}$.

7.2.1 Kollisjonsstrategier

Når flere signaler tidsmultiplekseres på samme buss, er det alltid en fare for kollisjoner. Det vil si at to eller flere pulser opptrer samtidig på inngangen til kretsen som er ansvarlig for å kode nervepulsene som adresse-hendelse par. Slike tilfeller vil være svært sjeldne i kameraet, siden det er maksimalt to nevroner som kan sende ut pulser samtidig som følge av at nevronene er koplet til to winner-take-all kretser. Det er også en egen instilling for kameraet som bestemmer minste opphold mellom to pulser. Dersom dette oppholdet varer f.eks. 100 ganger så lenge som varigheten av en puls, vil sannsynligheten for en kollisjon være én prosent. Selv om sannsynligheten er svært liten, må enhver AER krets likevel ha en kollisjonsstrategi. Det er hovedsaklig tre forskjellige prinsipper for kollisjonshåndtering:

Full arbitrering. Nevroner sender en forespørsel om tilgang til bussen før de legger ut sin adresse på den. Hvis bussen er opptatt, blir de satt på vent. Fordelen med denne strategien er at ingen pulser går tapt. Ulempen er at pulser kan bli forsinket.

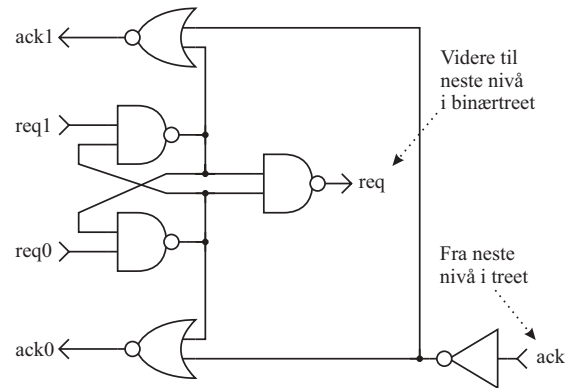
Forkasting. Denne strategien kontrollerer ikke tilgang til bussen. Hvis flere nevroner sender samtidig på bussen, vil det bli en kollisjon. Adressene til nevronene er ikke plukket tilfeldig. De er tildelt på en slik måte at en kollisjon kan detekteres av mottakeren. Dersom mottakeren oppdager at en kollisjon har funnet sted, forkastes pulsen, som selvfølgelig kan være en ulempe. Fordelen er at ingen nevroner blir forsinket og at kretsen er enkel.

Foreldelse vs. tap strategien. Dette er en kombinasjon av de to ovenstående. Her vil pulser som kolliderer bli satt på vent (bufret) i tilfelle kollisjon, men dersom en puls blir for mye forsinket blir den forkastet. Tanken er at man ikke ønsker å forkaste pulser, men at veldig forsinkede pulser kan gjøre mer skade enn godt, slik at det er bedre å forkaste disse. Ulempene er en treg og komplisert krets.

Kameraets adresse-hendelse kretser har full arbitrering som sin kollisjonsstrategi. Bakgrunnen for dette valget er at veileder allerede hadde disse kretsene ferdig laget. Siden oppgaven ikke har hovedvekt på adresse-hendelse representasjon, var det greit å benytte ferdige kretser som var testet tidligere. Jeg går derfor nærmere inn på full arbitrering strategien, men utelater en del detaljer.

I full arbitrering tar senderen seg av kollisjonshåndtering. Nevroner som ønsker å sende ut en puls sender en forespørsel til en utvelger (arbiter) -- en krets som velger ut et av flere signaler. En enkel utvelger er en som bare har to innganger, og velger en av disse. For å lage en krets som har flere innganger, kan man bruke kretsen som byggeblokk i et binærtre av mange kretser. Forespørsel om buss-tilgang går da flere nivåer ned til roten av treet før tilgang gis til det utvalgte nevronet. Signalet om at

nevronet kan få sende på bussen sendes da tilbake gjennom binærtreet. En utvelger med to innganger er vist i figuren nedenfor.



Dersom flere pulser kommer samtidig, vil de sendes ut etter hverandre. Digital logikk fører pulsen steg for steg mot roten i et binærtre. Så fremt pulsrate (antall pulser per sekund) ikke overstiger hastigheten til kretsene utenfor brikken, vil alle pulser bli sendt ut, om nødvendig køet underveis. Unntaket er i tilfeller der man får en ny puls på samme inngang før den første er sendt ut. Dette har med hvordan kømekanismen er laget.

7.3 AER som kommunikasjonsprotokoll for systemet

Kameraet har to winner-take-all kretser, med hhv. 43 og 52 utganger (42 rader og 51 kolonner pluss to inaktivitetsnevroner). Når en vinner-puls genereres, skal denne sendes ut av brikken. Slik gir kameraet beskjed om (rad, kolonne) koordinatene der et objekt i bevegelse befinner seg. Alternativt kan det sende ut pulser fra inaktivitetsnevronene, som beskrevet over.

Kameraet har en 100-pinnars innkapslingsbrikke, hvilket er ganske stort. Av økonomiske grunner har silisiumbrikken med kretsen 84 padder istedenfor 100. Dette sparer areal og blir derfor billigere. Det naturlige ville selvfølgelig vært og innkapslet brikken i en 84 pinnars kapsel, men vi har valgt en 100 pinnars innkapsling for at brikken skal passe på et kretskort vi har ferdig laget til et annet prosjekt (CAVIAR prosjektet).

De totalt 95 utgangene fra winner-take-all kretsene for kolonnene og radene er koplet inn på en AER krets. Når denne kretsen mottar en puls fra en av inngangene sender den ut adressen til inngangen på utgangsbussen. Fordi adressen kodes binært, trenger bussen på utgangen av kretsen 7 ledninger, som gir oss $2^7 = 128$ mulige adresser. Dette holder fint for kameraets 95 adresser.

”Address event representation” kretsen er delt opp i to seksjoner, en som er tilkopleet radene og en for kolonnene. Den tilordner radene adresser fra 0 til 42 og kolonnene adresser fra 64 og oppover. Dermed er det lett å se på den mest signifikante biten om adressen tilhører en rad eller en kolonne.

Det finnes flere kollisjonsstrategier, som beskrevet. Den spesifikke AER kretsen som kameraet benytter har såkalt ”full arbitration”, *og er laget av veileder.*

Tiden fra en puls er sendt ut til det kan komme en ny en, altså maksimal pulsrate, kan enkelt begrenses i kameraet ved å justere strømmen hver endringspiksel trekker, og også rad- og kolonnestrømmene. Videre kan hvileperioden i winner-take-all kretsen settes lang for å begrense pulsraten. Faren for å overkjøre en mottaker eller å sende en ny puls på en og samme inngang før den forrige er prosessert, kan dermed gjøres minimal.

7.4 Målinger på hele kretsen

Nå som både endringspikslene og winner-take-all kretsene er testet, gjenstår å se hvordan det hele fungerer i sammenheng. De største forskjellene mellom testkretsene og hovedkretsen er at

- ▶ Hovedkretsen har flere endringspiksler: tidligere er en enkelt endringspiksel testet, nå testes en matrise av 42x51 tettpakkede pikseler
- ▶ Hovedkretsen er større winner-take-all kretser: tidligere tester er i hovedsak utført på et fire-nevrøns WTA. Nå testes en 43 nevrons og 52 nevrons WTA.
- ▶ Tidligere har man testet hvor mye strøm en enkelt endringspiksel ønsker å trekke, nå tilkoples hver endringspiksel en rad og en kolonne, som er tilkoplest en strømkilde, som må stilles inn.
- ▶ Vi kan ikke lenger måle V_{soma} for det vinnende nevronet like enkelt som før. Noen nevrons i hovedkretsen har målepunkter, men det er vanskelig å få akkurat disse til å vinne, bortsett fra "timeout" nevronsene som kan stimuleres til å vinne, fordi de er kontrollert.
- ▶ Adresse hendelse protokollen kommer inn i bildet: Utgangspulsene fra winner-take-all kretsen leses av som adresse-hendelse par som mottas av et interface kort i en PC. Naturlig nok må et program benyttes til å kontrollere utlesningen av pulser, og dette fører til forskjeller fra testkretsene, og mulige feilkilder.

Jeg har gjort små endringer i et program som leser AER output fra et interface kort i en PC. Programmet skriver nå ut innkommende hendelser på skjermen (i konsoll) så snart de kommer, sammen med en enkel indikasjon (tekst "grafikk") på hvilken rad eller kolonne hendelsen representerer. Øyeblikkelig tilbakemelding mens man beveger noe foran linsen eller justerer bias spenninger er praktisk. Et eksempel er at man kan følge med på om man flytter objektet utenfor synsfeltet til kameraet. Det hjelper også for innstillingen av kretsen at man kan se om det kommer mange hendelser.

Som tidligere nevnt for lysfølsomme kretser, må testene utføres med alt kunstig lys (lysstoffrør spesielt) som er tilkoplest strømmettet. Lommelykt er greit å bruke, da den går på batterier og ikke forstyrrer på samme måte.

Inaktivitetsnevronsene ble stilt inn ved å følge med på deres V_{soma} på oscilloskopet og justere inngangsstrømmene.

7.4.1 Virkningen av den uønskede spenningsforskjellen V_{out} - $V_{feedback}$

Den uønskede spenningsforskjellen mellom V_{out} og $V_{feedback}$ i fotokretsen, gjør at kameraet ikke ser endringer slik det var ment å gjøre. Snarere kan det følge et lyspunkt rundt i synsfeltet. Forskjellen på ønsket og faktisk virkemåte er at når lyspunktet stopper å bevege seg, skulle også kretsen slutte å sende ut pulser. Det gjør den ikke. Så lenge det er økt lys vil den sende ut hendelser. Likevel er resultatet ikke så dårlig.

7.4.2 Generelt om resultatene

For lav pulsrate gjør responsen fra kretsen treg. Når et objekt beveges over synsfeltet, rekker bare noen få pulser å bli sendt ut, og kretsen følger derfor ikke objektet jevnt. Dette rettes opp ved å øke pulsraten (ved å øke rad- og kolonnestrømmene og strømmen hver endringspiksel trekker).

Selv med større pulsrate opplevdes utgangene fra kretsen å være litt hakkete. Med det mener jeg at en jevn bevegelse fra den ene siden til den andre ikke resulterer i at en og en rad/kolonne ”pent og pyntelig” vinner, men snarere at noen hoppes over. Det ble eksperimentert mye med å justere de enkelte strømmene, pulslengden, lengden av hvileperioden, inaktivitetsnevronenes inngangsstrøm og liknende, men helt jevnt ble det aldri.

Årsaken til at det ikke virker mulig å få alle radene/kolonnene til å vinne en etter en ved en jevn bevegelse, er trolig en kombinasjon av følgende:

- ▶ Noen nevroner er sterkere enn andre, og har lettere for å vinne. Det kan være at kondensatorene er mindre (lades opp raskere) eller at forsterkerens terskelspenning er lavere (sender ut en puls ved lavere spenning).
- ▶ Rad/kolonne strømkildene (p-transistorer) kan ha relative forskjeller, men disse er såpass store ($33\ \mu\text{m} \times 12\ \mu\text{m}$), at forskjellene ikke skulle være alt for store. Dette kan likevel føre til at noen rader/kolonner er sterkere enn sine naboer, og vil vinne dersom stimuli til disse radene er noenlunde likt (de observerer like mye endring).
- ▶ Detaljer i implementasjonen av adresse-hendelse protokollen kan også være en årsak til at kretsen noen ganger hopper over nabo rader/kolonner når man beveger objektet jevnt over dem. Dette fordi at når en rad/kolonne venter på å sende ut sin puls, sperrer den raden/kolonnen ved siden av pga. trestrukturen.

7.4.3 Forsøk som ble gjort

Flere ”objekter” ble forsøkt beveget i synsfeltet til linsen for å stimulere kretsen. De som virket best var de som reflekterte lyset. Lys rett inn i linsen fungerte også, men man måtte da hele tiden vinkle lommelykten slik at den lyste mot linsen mens man flyttet den, hvis ikke ble lyset mye svakere. Det som virket best var:

- ▶ ”lysende finger” – lommelykt lyser på en fingertupp
- ▶ ”lommelykt gjennom oransje papir” – et tykt oransje papir holdes over linsen (ca. 20 cm fra) mens lommelykten holdes inntil papiret og beveges

- refleksjon fra taket – lommelykten lyser i taket, og kameraet følger refleksjonen

Følgende tester ble utført med ”lysende finger”.

1. En jevn bevegelse fra den ene kanten av radene til den andre. Kolonnene var avslått.
2. En jevn bevegelse fra den ene kanten av kolonnene til den andre. Radene var urørt ($V_{timeoutRAD}$ og $V_{biasRAD}$ som i 1).
3. Bevegelse frem og tilbake over radene. Ingen endringer i innstillinger fra 2.
4. Bevegelse frem og tilbake over kolonnene.
5. Bevegelse frem og tilbake langs den ene diagonalen.
6. Bevegelse frem og tilbake langs den andre diagonalen.

Testene ble repetert med ”lommelykt gjennom oransje papir”. I tillegg ble følgende tester utført:

7. Bevegelse i firkant.
8. Bevegelse i sirkel.

Disse to testene, nr 7 og 8, var vanskeligere å både utføre og kontrollere resultatet av enn tidligere tester. Dette kommer særlig av at kameraet sender ut enten en kolonne puls eller en rad puls, ikke (rad, kolonne) par.

7.4.4 Rad og kolonne par

Bevegelser som stimulerer både radene og kolonnene, som diagonaler, sirkeler og firkanter, trenger verdi for både rad og kolonne for å kunne vite hvor objektet er. Siden kameraet kun sender ut enten rad eller kolonne, må slike (rad, kolonne) par settes sammen etterpå. Det gjøres ved å kombinere siste verdi for kolonnene og radene til par. Det sier seg selv, at dersom det går lang tid uten en puls for enten radene eller kolonnene, kan (rad, kolonne) kombinasjonen har en feil verdi for enten raden eller kolonnen (unntatt hvis raden/kolonnen ikke er endret).

Tester som stimulerer både radene og kolonnene er avhengige av at begge disse sender ut pulser jevnlig. For bevegelse over diagonalene vil det optimale være om radene og kolonnene sender ut pulser omtrent hver annen gang. Da vil (rad, kolonne) parene bli riktige.

Det viser seg at radene og kolonnene trolig forstyrrer hverandre, selv om de skal være uavhengige, da de har hver sin winner-take-all krets. Likevel observerte jeg at særlig kolonnene i mange tilfeller stopper å sende ut pulser når radene sender ut sine pulser (bl.a. ved bevegelse over diagonal). Kombinasjonen av rad- og kolonner blir derfor svært dårlig i enkelte tilfeller.

Man kan merke seg at rad og kolonne par ikke er nødvendig for at signalene skal kunne brukes av utstyr utenfra, kun for plotting. Tilkoplet utstyr kan for hver utgangspuls, enten det er en rad eller en kolonne, snu seg tilsvarende, uten å endre den andre aksen. Rad og kolonne par blir derfor mest for å vise i plott.

7.4.5 Forsøk på å forbedre sirkel- og firkant resultat

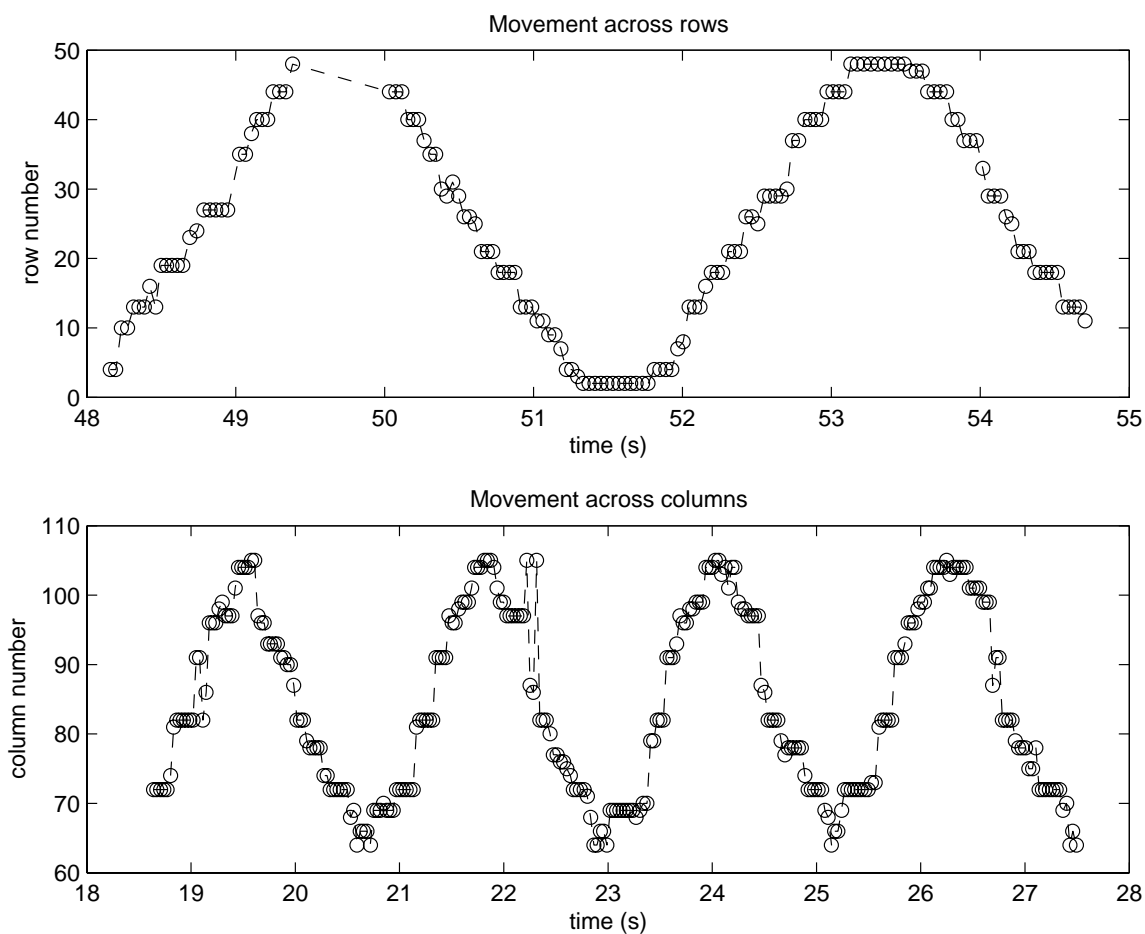
Resultatet av test nr. 7 og 8 ble forsøkt forbedret ved å øke antall pulser. Tanken er at det vil alltid være en del målepunkter som faller utenfor hovedgruppen. Et histogram vil da tydelig få frem firkant/sirkel bevegelsen som er gjort.

Det viste seg nå at Windows konsollen ikke klarer å holde følge med den høye puls-raten. I noen eksperimenter kunne konsollen vise 10 pulser mens programmet registrerte 1000. Forsøk ble gjort der utskrift til konsoll ble slått av eller ble satt til å vise hver n-te puls, men dette gjorde ikke resultatene noe bedre. Alle pulser lagres i buffer på kortet, og kommer derfor med i filen, så det er lite sannsynlig at utskrift til konsoll forstyrrer resultatene.

Problemet med kombinasjon av rader og kolonner til par, gjorde at målingene for bevegelse i sirkel og firkant heller ikke nå ble særlig gode. Disse er derfor ikke plottet.

7.4.6 Bevegelse langs radene eller kolonnene

Bevegelse langs enten radene eller kolonnene, der pulser for kolonner er fjernet i rad-målingen og omvendt, er vist i Plott 14.



Plott 14 - over: bevegelse langs radene, under: bevegelse langs kolonnene

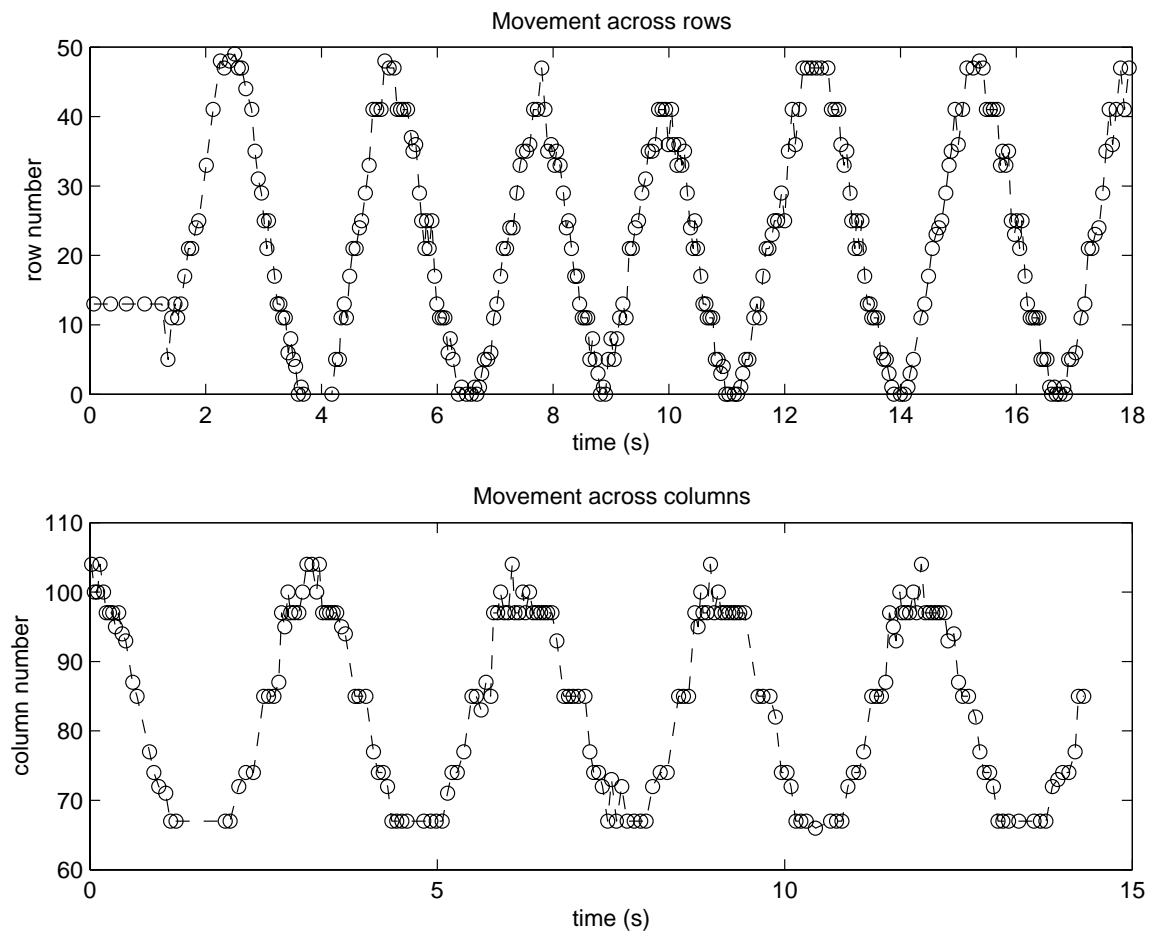
7.4.7 Problem med kollisjoner?

Som nevnt i avsnitt 7.4.4 kan det virke som om kollisjoner mellom radene og kolonnene er et problem. Dette er på grunnlag av observasjoner der man forventer aktivitet fra både radene og kolonnene (bevegelse langs diagonalene er et eksempel) og ser at når radene sender ut pulser forholder kolonnene seg i ro og omvendt.

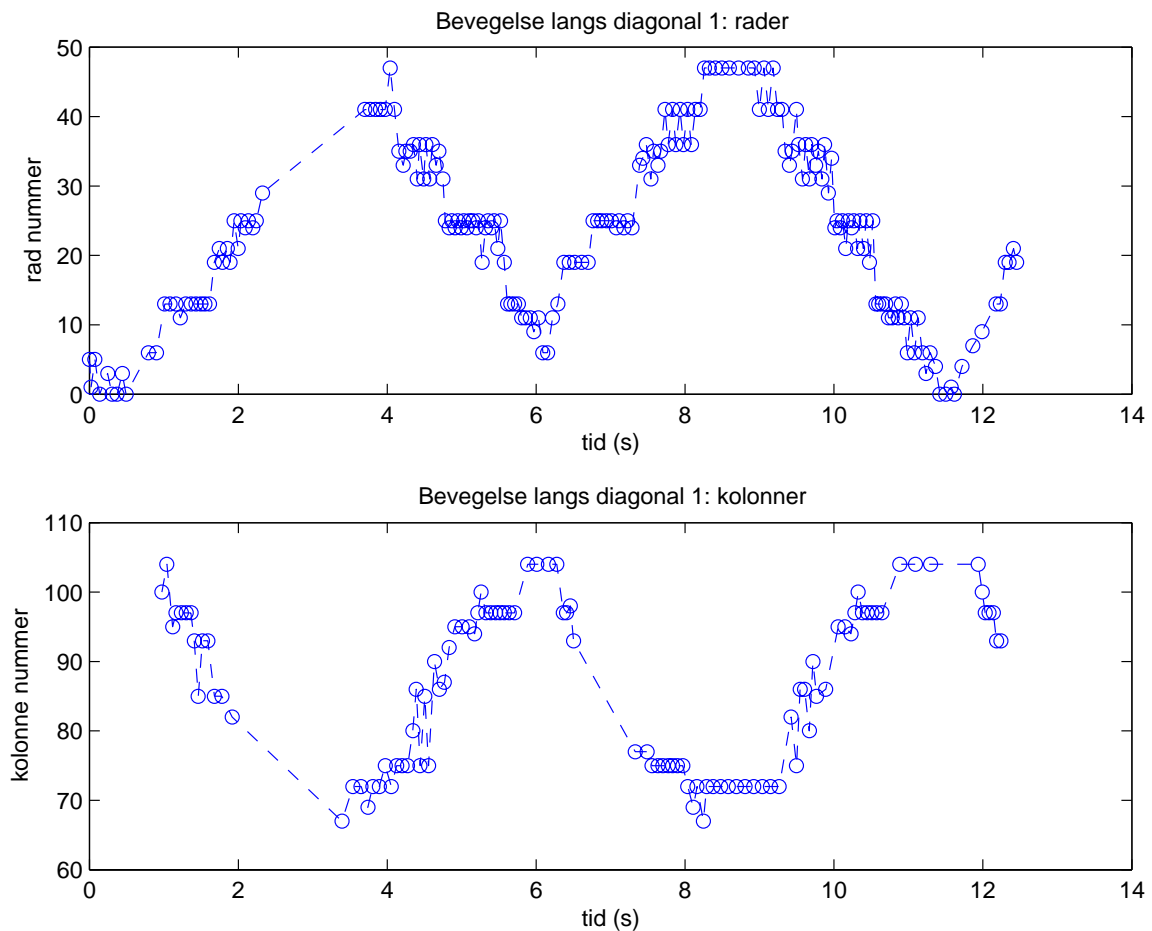
Radene og kolonnene er som nevnt tilkoplek hver sin WTA krets, og sender ut hendelser uavhengig av hverandre. De to WTA kretsene er likevel tilkoplek *samme* AER krets, så helt uavhengige er de likevel ikke. Hvis de to WTA kretsene sender ut en puls samtidig, får vi dermed en kollisjon.

Pulsraten er forsøkt skrudd ned, for å unngå mulige kollisjoner, på samme måte som forklart i avsnitt 7.4.2. Underveis i justeringen observerte jeg frekvensen av pulser. For å se alle pulser som sendes ut fra samtlige nevroner kan man se på radene og kolonnenes *reset* linjer, som er høye hver gang enten radene eller kolonnene sender ut en puls.

Resultatene er vist i plottene som følger.



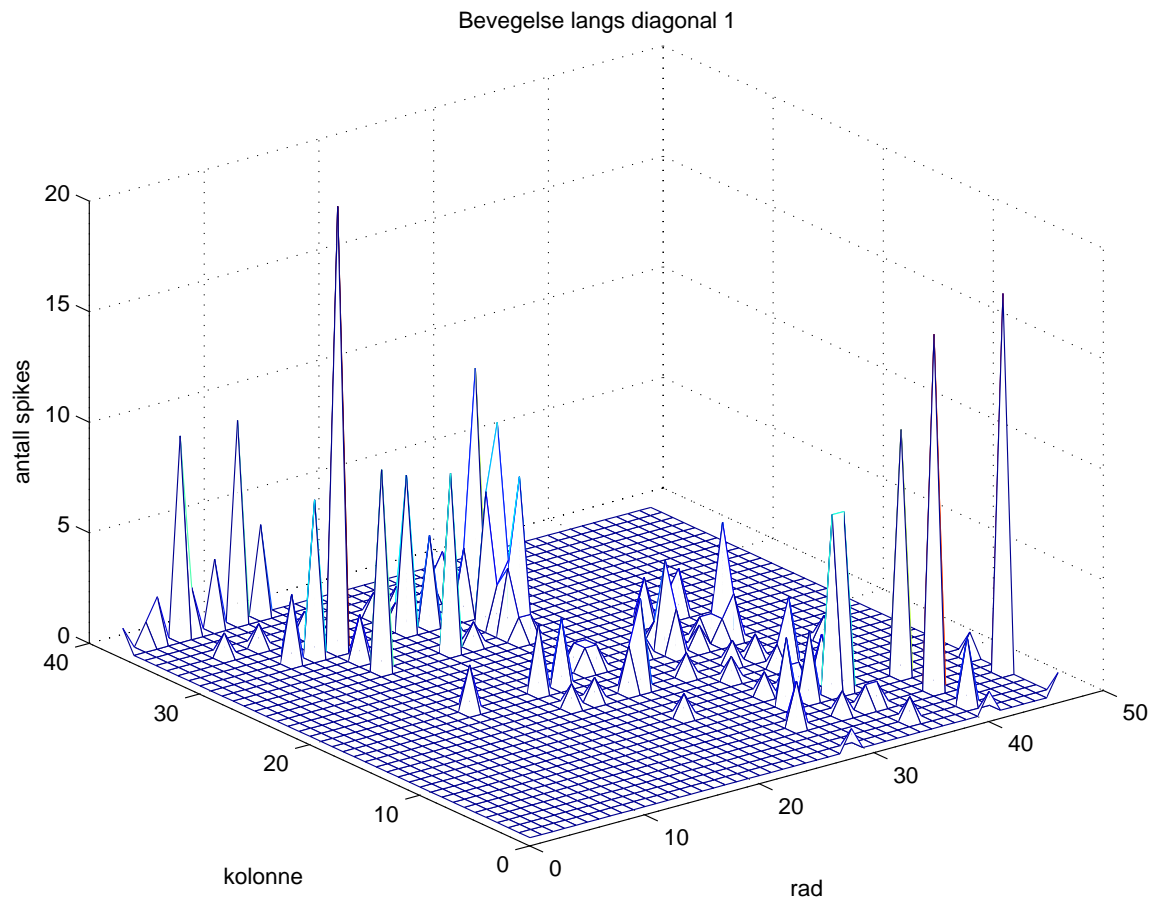
Plott 15 - bevegelse langs radene og kolonnene (hvert sitt plott). Gjennomsnittlig pulsfrekvens er 16,7 Hz i øvre plott og 20 Hz i nedre plott.



Plott 16 - bevegelse langs diagonal 1. Radene (øverst) og kolonnene (nederst) skal bevege seg omvendt av hverandre.

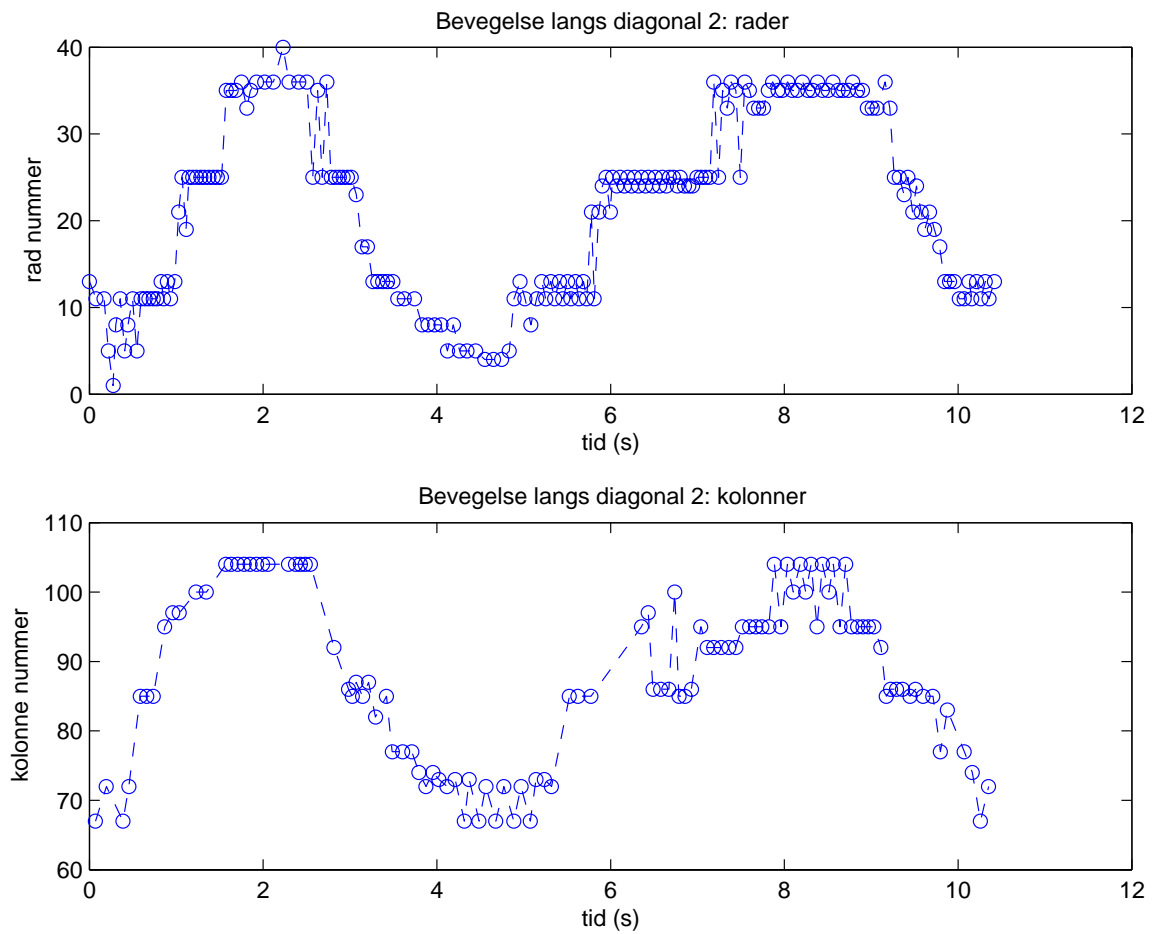
Som man kan se i Plott 16 forekommer det perioder der ingen pulser blir sendt ut fra enten radene eller kolonnene. For eksempel sender kolonnene ikke ut pulser i tidsrommet $t = 2$ s til $t = 3,5$ s. Alle parene som genereres for innkomne pulser fra radene i dette tidsrommet vil ha samme verdi for kolonnene, som er kolonne nr. 82.

Antall forekomster av hvert (rad, kolonne) par er så summert opp, og vist i histogrammet nedenfor.



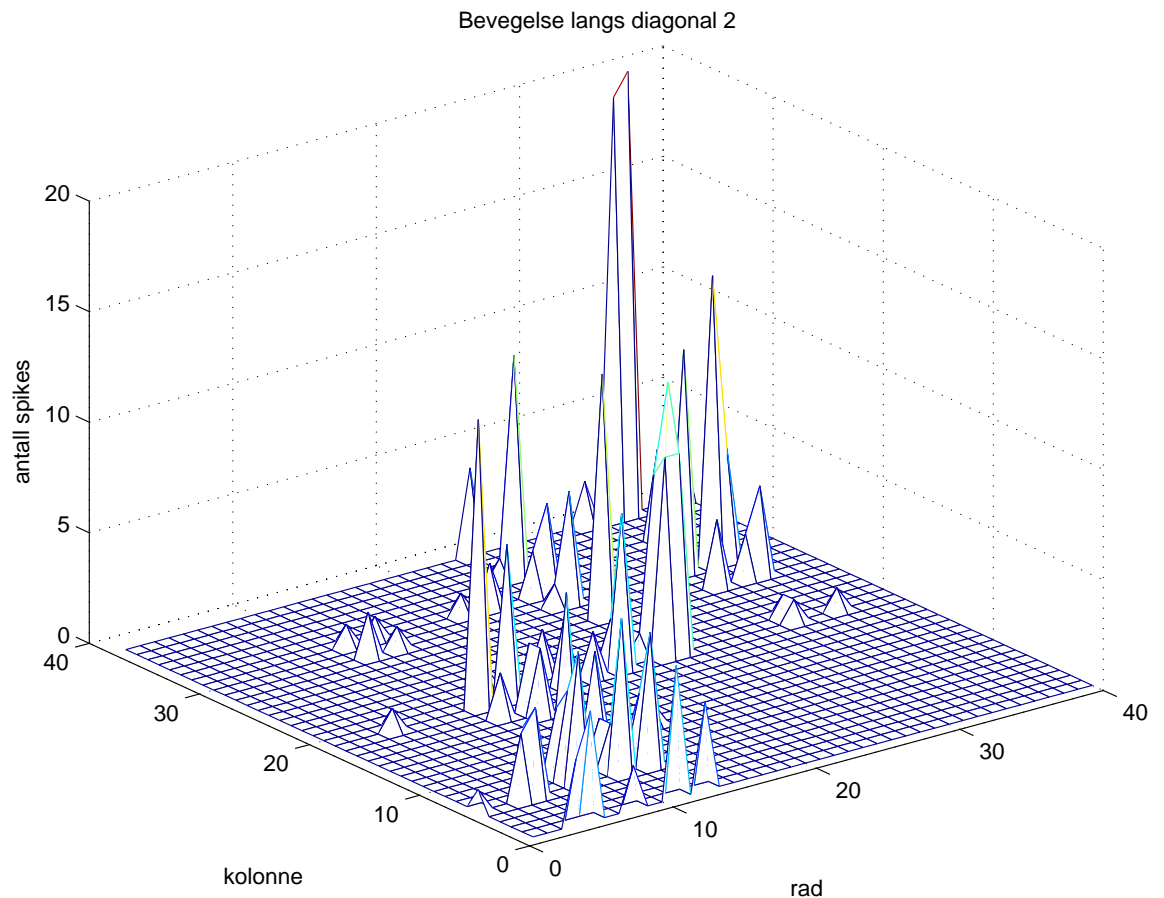
Plott 17 - bevegelse langs diagonal 1. Histogram viser antall forekomster av hvert kombinerte koordinatpar. Gjennomsnittlig pulsfrekvens var 24 Hz.

Tilsvarende målinger som for diagonal 1 er her gjort for diagonal 2. Vi vil nå se at målingene fra radene og kolonnene beveger seg parallelt, de vil gå mot høye og lave verdier samtidig. Det ene hjørnet har koordinatene (rad nr. 0, kolonne nr. 60) og det andre (rad nr. 41, kolonne nr. 110).



Plott 18 - bevegelse langs diagonal 2. Radene (øverst) og kolonnene (nederst) skal endre seg parallelt.

Igen er pulsene kombinert fra radene og kolonnene til (rad, kolonne) koordinater. Dette er vist i histogrammet på neste side.



Plott 19 - bevegelse langs diagonal 2. Antall forekomster av hvert genererte (rad, kolonne) par er vist som et histogram. Gjennomsnittlig pulsfrekvens var 29 Hz.

Måling av sirkelformet bevegelse er vanskelig å vise som vanlig plott (radene og kolonnene hver for seg), da ikke er lett å se noen sammenheng mellom rad- og kolonne målingene. Kombinert som histogram er beste måte å vise resultatet på, men dette gir dårlige resultater, pga. rad- og kolonne kombinasjonen til par.

Bevegelse i firkant gir også mindre gode resultater, av samme grunner som for sirkelbevegelse. Gjennomsnittlig pulsfrekvens ved sirkel- og firkantbevegelse var henholdsvis 17 Hz og 16 Hz, som er lavere enn for eksempel bevegelse langs radene og kolonnene. Dette kommer sannsynligvis av at objektet ofte beveges utenfor synsfeltet til kameraet, da jeg forsøkte å lage så store sirkler og firkanter som mulig (bevege langs kanten av synsfeltet).

8 OPPSUMMERING OG VIDERE ARBEID

8.1 Kommentar av resultatene

Slik jeg ser det fungerer kameraet meget bra, bortsett fra at det ulineære elementet ble plassert feil vei. Dette ga den uønskede avstanden mellom gammel og ny verdi for lyset som igjen gjorde at endringer ikke ble detektert slik som forventet.

Det finnes likevel noe positivt ved dette – at feilsøking kunne finne sted. At feilen ble identifisert og verifisert ved simuleringer er en god erfaring å ha. Det ble også lagt en del arbeid i å analysere virkningene av feilen, som bl.a. at endringer til lysere og mørkere tilpasses med ulik hastighet.

8.2 Hva har arbeidet bestått i

Under arbeidet med kretsen har jeg gjennomgått en hel syklus, fra idé via studier av biologi og kretser, kretskonstruksjon, simuleringer, fysisk utlegg og testing på laboratoriet tilbake til ideer for videre utvikling. Jeg har også måttet skrive og dokumentere alt som er gjort, og lese om hva andre har gjort, samt sette meg inn i hvordan et kamera fungerer.

Winner-take-all kretsen som ble utviklet sammen med kameraet virker svært bra, og er innsendt til konferansen 2004 IEEE International Symposium on Circuits and Systems. Arbeidet med å skrive en artikkel til en konferanse har vært svært lærerikt.

8.3 Utvidelser / videre arbeid

Det hadde vært fristende å rette opp feilen med det ulineære elementet, altså å sette det riktig vei, og se om resultatene nå blir mer som forventet. Da det er svært kostbart å sende kretser til produksjon, er det ikke aktuelt. Det er likevel nye prosjekter på gang, med nye hovedfagsstudenter, som vil inkludere endringspikslene fra dette prosjektet, om noe modifisert.

8.3.1 Flere objekters følger

Som diskutert i avsnitt 3.2.1 hadde det vært en utfordring å lage en krets som kan følge flere objekter. Dette er såpass krevende at det burde holde som en doktorgrads oppgave tror jeg. Det er mange forskjeller fra denne oppgaven.

8.3.2 Winner-take-all med hysteresis

En winner-take-all krets med hysteresis, altså at en vinner får fordeler slik at kretsen ikke så lett skifter vinnere, er en idé jeg har sett brukt. Det kan hjelpe hvis kretsen har vanskelig med å holde fast ved et objekt (blir lett distraheret med andre ord). Andre utvidelser av winner-take-all kretsen er diskutert i ”paperet” vedlagt.

8.3.3 Rad og kolonne annenhver gang

På grunn av problemene med kombinasjon av rad- og kolonneverdier til par, kunne en tenkt utvidelse være å endre på kretsen slik at den ble tvunget til å sende ut rad og

kolonne annenhver gang. Dette burde ikke være alt for krevende å lage, da man lett kan tenke seg at de to WTA kretsene kunne ha en hvileperiode som var veldig lang, der man venter på at den andre WTA kretsen sender ut sin puls. Når WTA kretsen for radene sender ut en puls vil den avbryte hvileperioden til kolonnenes WTA og omvendt. Dersom den ene av kretse ikke sender ut en puls, fordi det ikke er noen endringer langs dens akse, vil hvileperioden uansett avsluttes etter en bestemt tid.

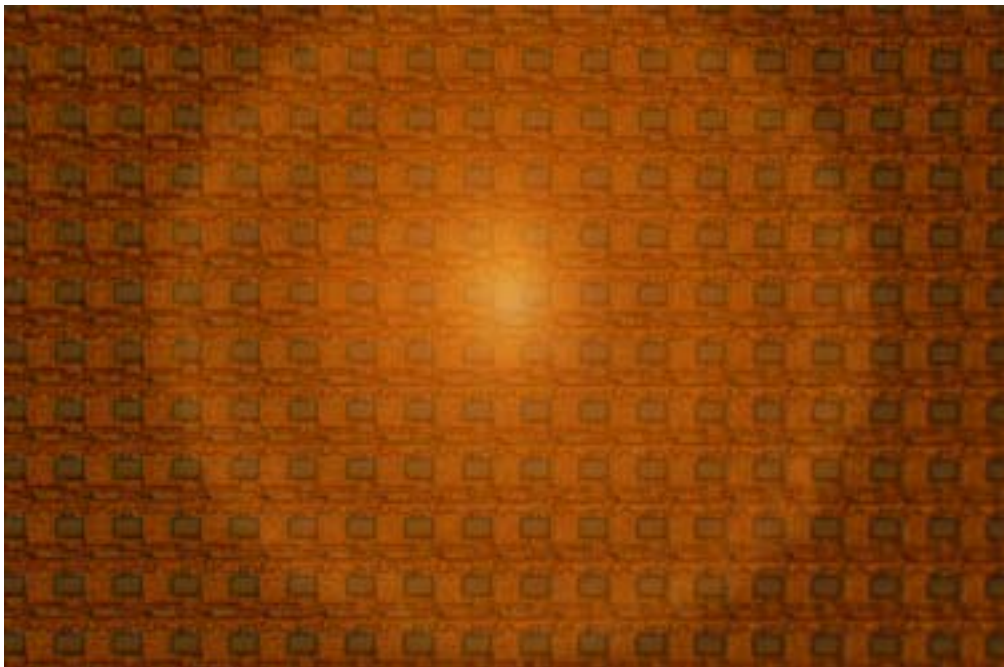
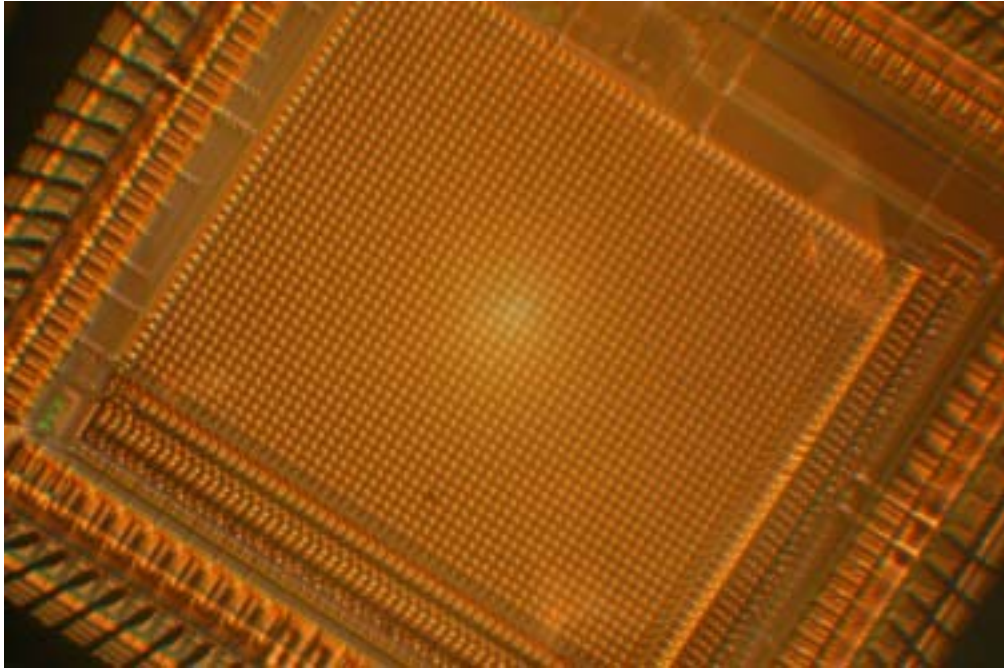
9 LITTERATUR

Bibliografi

- Bohaen, K.A. "A retinomorphic vision system," *IEEE Micro* (16:5) 1996, pp 30-39.
- Bohaen, K.A. "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Transactions on circuits and systems* (47:5) 2000, pp 416-434.
- Cariani, P.A. "Temporal coding of sensory information.," in: *Computational Neuroscience: Trends in Research*, J.M. Bower (ed.), Plenum, Boston, Massachusetts, 1997, pp. 591-598.
- Delbruck, T. "'Bump' circuits for computing similarity and dissimilarity of analog voltages," International Joint Conference on Neural Networks, Seattle Washington, 1991, pp. 475-479.
- Delbruck, T., and Mead, C. "Adaptive Photoreceptor with Wide Dynamic Range," International Circuits and Systems Meeting, London, 1994.
- Kolb, H. "How the Retina Works," in: *American Scientist*, 2003.
- Lazzaro, J., Ryckebusch, S., Mahowald, M.A., and Mead, C. "Winner-take-all networks of O(n) complexity," *Advances in Neural Information Processing Systems* (1) 1988, pp 703-711.
- Mead, C. *Analog VLSI and neural systems* Addison-Wesley, Reading, Mass., 1989, pp. 371 s., 378 pl.
- Moini, A. "Vision Chips or Seeing Silicon," The Centre for High Performance Integrated Technologies and Systems, Department of Electrical & Electronics Engineering, The University of Adelaide, Australia.
- Serrano, T., and Linares-Barranco, B. "The Active-Input Regulated-Cascode Current Mirror," *IEEE Transactions on circuits and systems* (41:No. 6), June 1994 1994, pp 464-467.
- Vittoz, E.A. "Analog VLSI Signal Processing: Why, Where and How?," in: *Analog Integrated Circuits and Signal Processing*, 1994, pp. 27-44.

10 VEDLEGG

10.1 Foto av kretsen



10.2 Linsen og linseholderen

For prosjektet ble det kjøpt inn en vidvinkel linse (brennvidde 4 mm, ikke sammenliknbart med kamera linse for vanlig 35mm film) som vanligvis brukes i overvåkningskameraer. Denne linsen er standard CS montert, som betyr at avstanden fra bakstykket av linsen til brennpunktet er 12,5 mm. Linsens nedslagsfelt følger 1/3" standarden, som vil si at mye lys faller utenfor sensoren som bare er på ca. 2x2 mm. Kameraet får dermed ikke på langt nær så vid vinkel som linsen tilsier (76,9° diagonalt over sensoren) men ca. en firedel (19°). Linsen kan fokusere så nær som 20 cm, og har justerbar blender.

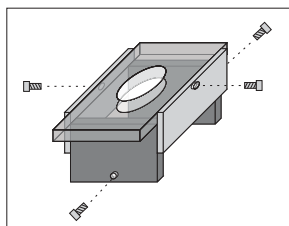
For å montere linsen på chipen var det nødvendig å designe en linseholder. Denne linseholderen måtte passe til en innkjøpt ZIF 100 sokkel som brukes i CAVIAR prosjektet. Denne sokkelen har en brysom arm som brukes til å feste brikken med. Til forskjell fra andre sokler, der denne armen kan legges helt ned parallelt med kretskortet, står denne armen alltid opp, og gjorde det vanskelig å lage en god linseholder.

Kravene til linseholderen er, i tillegg til å holde linsen på plass:

- ▶ Avstanden fra bakstykket av linsen til brikken må være nøyaktig 12,5 mm slik at bildet blir fokusert.
- ▶ Så lite strølys som mulig må slippe til. Dette er blant annet oppnådd ved å lakkere linseholderen sort på innsiden.
- ▶ Objektivets posisjon bør kunne justeres i x- og y-retning siden sensorene ikke alltid er plassert midt på brikken.

De tekniske tegningene for linseholderen er vist på neste side. Linseholderen ble laget ved verkstedet på fysisk institutt ved Universitetet i Oslo.

Objektivets merke/modell er Computar T0412FICS-3.

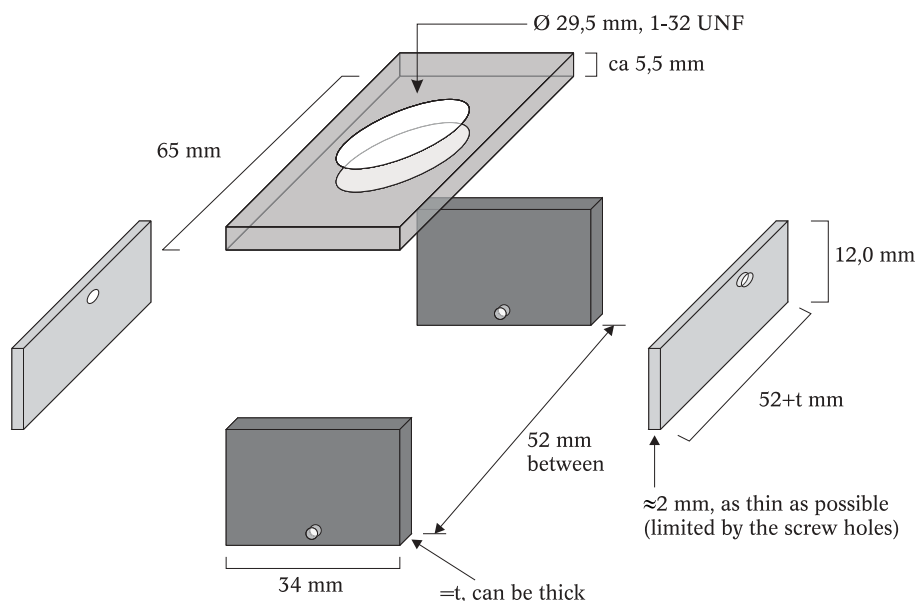
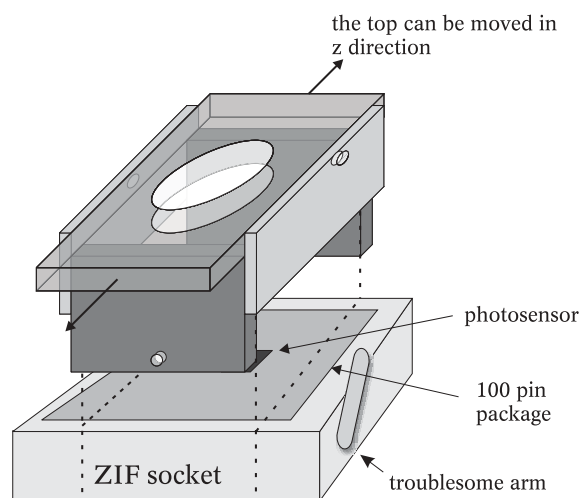


Lens holder for 100 pin ZIF socket

NB: The top plate is transparent for illustrative purposes only. In reality it will NOT be transparent

Front and back on the outside of the ZIF socket.
Left and right sides rests on top of the chip package.

whole lens holder can be moved in x direction



Critical measurments: the height of the left and right sides must be 12.0 mm for the lens to focus correctly (CS mount lens).

Additional info: The left and right sides can be fastened to the front and back any way practial, e.g. with screws. If screws are used, the front and back plate must be thick enough for screw holes to fit.

Linseholderen lakkeres matt sort på innsiden.

Jens Petter Abrahamsen, jenspa@ifi.uio.no, tlf. +47 952 03 663, august 2003

10.3 Målepunkter, testkretser og innstilling av kretsen

På grunn av den store tettheten i selve matrisen av endringspiksler (endringsensitive punkter) er det ikke mulig å få målt spenningen i punkter inni i matrisen. Dette er først og fremst fordi det er vanskelig å strekke ledninger inn til et punkt midt inne i matrisen. Et alternativ er å måle på punkter i utkanten av matrisen, men også dette kan være risikabelt med hensyn på forstyrrelser. Som ved alle målinger, er det viktig at målingen ikke forstyrrer originalen. Da vil jo målingen vise feil og være verdiløs i tillegg til at originalen er ødelagt.

Alle målinger av spenninger må derfor være kopier av originalen. For å lage en kopi av en spenning bruker vi en spenningsfølger. En best mulig kopi får vi med en "wide-range" følger, som kan kopiere spenninger som er nær jord og forsyningsspenningen. Denne spenningsfølgeren bør plasseres så nær det punktet vi skal kopiere spenningen i som mulig. En lang kabel fra et punkt kan bidra til en stor parasittkapasitans i punktet. Siden punktene konkurrerer i et "winner-take-all" nettverk, er det lite ønskelig at noen punkter blir hemmet på denne måten.

10.3.1 Målepunkter i hovedkretsen

De endringspikslene som det måles på i matrisen er derfor de helt i hjørnene. Der er det mulig å trekke korte ledninger for å kople til spenningsfølgere. De punktene i kretsen som det skal måles på er de som jeg har målt på simuleringene. For enkelthets skyld måler jeg kun spenninger, selv om jeg på simuleringene også har målt strøm. Følgende testpunkter finnes i hovedkretsen, hvorav de fire første er essensielle for innstilling av kretsen som beskrevet senere:

- ▶ Spenningen på rad nr. 40, altså linjen som alle endringspikslene i rad nr. 40 er tilkoplethet. (V_{rad40})
- ▶ For sikkerhets skyld spenningen som over, men på en annen rad, rad nr. 41. (V_{rad41})
- ▶ Spenningen på kolonne nr. 0 (V_{kol0})
- ▶ Spenningen på kolonne 1, for sikkerhets skyld. (V_{kol1})

Fra de to "time-domain winner-take-all" kretsene i hovedkretsen tar jeg ut følgende målepunkter:

- ▶ Spenningen V_{soma} for nevron nr. 42. Det er det nevronet som får sin inngangsstrøm fra rad nr. 42. Denne spenningen viser hvor langt nevronet er ladet opp og om det sender ut pulser. Det kan være greit å måle dette før winner-take-all kretsen koples inn. ($V_{nevro42Kopi}$)
- ▶ Som over, men for inaktivitetsnevronet for radene. Det er alltid interessant å ha en egen måling av inaktivitetsnevronet. ($V_{nevroInactiveKopi}$)
- ▶ Som over, men nå V_{soma} for nevron 0 tilkoplethet kolonne nr. 0. ($V_{nevro0Kopi}$)
- ▶ Som over, men for kolonne nr. 1. ($V_{nevro1Kopi}$)
- ▶ Digital forespørsel om nullstilling fra et hvilket som helst av radene. Nevronene som er tilkoplethet en rad betegnes som tilkoplethet i Y retning. Når "winner-take-all" kretsen skal undersøkes er denne spenningen viktig å se på. ($V_{resetReqY}$)

- ▶ Digital nullstillingsspenning for alle nevronene tilkopleet en rad, altså Y retning. Nullstillingsspenningen er sammen med forespørsel om nullstilling, viktig å se på for "winner-take-all" kretsen. (V_{resetY})
- ▶ Digital forespørsel om nullstilling av nevronene fra et av nevronene i X retning, altså et nevro som er tilkopleet en kolonne. ($V_{resetReqX}$)
- ▶ Digital nullstillingsspenning for nevronene i X retning. (V_{resetX})

10.3.2 Målepunkter i spesielle testkretser

En del punkter er det likevel ikke mulig å få målt i matrisen på grunn av den høye pakketettheten. Jeg har derfor laget noen testkretser i tillegg til hovedkretsene i kameraet. Disse testkretsene består av enkle deler og noen større mer sammensatte deler av kretsene kameraet er satt sammen av. Testkretsene er mer luftig pakket enn kretsene i kameraet slik at det er lett å komme til. Ved at kretsene er oppdelt oppnår man muligheten til å styre input til del-kretser som vanligvis vil ha lys som input eller som får sin input fra utgangene av andre kretser. Dette gjør det enklere å karakterisere enkelte deler av kameraet, og kan være veldig nyttig dersom noe går galt. Følgende testkretser er laget i tillegg til selve kameraets hovedkretser:

- ▶ En enkelt endringspixel bestående av fotodiode, fotokrets og antikorrelasjonskrets slik som endringspixelene i hovedkretsene. Denne har sin egen rad/kolonne strøm. Den trenger bias-spenninger for fotokretsen og antikorrelasjonskretsen. Følgende spenninger kan måles i denne kretsen:
 - ▶ Spenningen over fotodioden, V_P .
 - ▶ Utgangsspenningen fra fotokretsen, V_{out}
 - ▶ Spenningen på tilbakekoplingen, $V_{feedback}$
 - ▶ Utgangsspenningen fra antikorrelasjonskretsen, $V_{bumpOut}$. Utgangen fra antikorrelasjonskretsen er egentlig en strøm, men spenningen kan også måles over transistorene i strømspeilet på utgangen. Spenning er enklere å måle, så derfor er denne løsningen valgt.
- ▶ En enkelt endringspixel der fotodioden er byttet ut med en transistor. Strømmen gjennom transistoren kan styres utenfra og skal erstatte diodestrømmen. Det er mye enklere å kontrollere en spenning enn å kontrollere lyset inn på en fotodiode, så denne kretsen gjør det enkelt å sjekke spenningene V_P , V_{out} og $V_{feedback}$ i forhold til en kontrollert fotostrøm. Vi kan da mye enklere måle øvre og nedre hastighet som endringspixelen kan registrere. Dette kan også beskrives som øvre og nedre kuttfrekvens i et båndpassfilter.
- ▶ En enkelt fotodiode med tilhørende fotokrets. Denne gir oss muligheten til å måle spenningene i punktene V_{out} og $V_{feedback}$ uten at de er tilkopleet antikorrelasjonskretsen. Det eneste som er tilkopleet er da spenningsfølgere for disse to.
- ▶ En enkel antikorrelasjonskrets. Den får sine to innganger, V_{plus} og V_{minus} , direkte fra inngangspaddene. Dette er nyttig når det skal lages en karakteristikk av antikorrelasjonskretsen.
- ▶ Følgende testkretser er laget for å ytterligere måle og karakterisere det nevromorfe winner-take-all nettverket:

- En enkel celle i WTA kretsen, tilkople sin egen $V_{resetReq}$ (forespørsel om nullstilling) linje og sin egen V_{reset} (nullstilling) linje. En celle består av et nevron med tilhørende kretser for tilkopling til en nevro-WTA krets. Fordelen av å ha en enkelt celle er at man enkelt kan observere virkemåten til et nevron uavhengig av andre nevroner.
- Fire celler sammenkoplet til et lite winner-take-all nettverk. Inngangsstrømmen til de fire kan stilles uavhengig av hverandre og V_{soma} for alle fire nevroner er tilgjengelig utenfra brikken. Pulsene som nevronene genererer er også tilgjengelig på utgangspinnene. De fire cellene deler naturligvis en felles nullstillings-forespørsel linje og en nullstillingslinje.

For å spare inngangspad-er er noen av testkretsene tilkople samme inngang. Siden testkretsene skal brukes og måles på hver for seg, har det ingen ting å si om spenningen inn på kretser som ikke brukes er feil for øyeblikket. I de fleste tilfeller er likevel kun lignende innganger koplet sammen. For eksempel er spenningen som styrer kolonnestrømmen til flere av testkretsene den samme.

10.3.3 Prosedyre for innstilling av kretsen

Innstilling av kretsen består, i tillegg til å innstille linsen, av å sette riktige forspenninger. Innstilling av linsen samt en nærmere analyse av de forskjellige forspenningene er gitt senere. Forspenningene som må settes for kretsen er som følger:

1. $V_{fotoBIAS}$ er spenningen på p-transistoren på utgangen av fotokretsen. En korrekt innstilling av denne er ikke veldig følsom, og vil ifølge simuleringer ligge sentrert rundt 4,3 V med et godtatt avvik på 200 mV. Egne testkretser kan benyttes til å studere fotokretsen i mer detalj.

2. $V_{bumpBIAS}$ er forspenningen som bestemmer halestrømmen til antikorrelasjonskretsen. Som med fotokretsens forspenning, kan også denne settes løst og er heller ikke kritisk. $V_{bumpBIAS}$ bestemmer for mye strøm hver endringspixel trekker. Simuleringene tilsier en spenning på $4,1V \pm 0,2V$.

3. Forspenningene $V_{biasKOL}$ og $V_{biasRAD}$ styrer strømkildene plassert på toppen av hver rad og kolonne, og må stilles inn svært nøyaktig. Innstillingen blir lettet av at det er målepunkter på to kolonner og to rader som vi kan ta til hjelp. Rad- og kolonnestrømmene må settes akkurat store nok til at alle endringspixelene kan trekke ønsket strøm, men ikke større. For å stille inn $V_{biasKOL}$ og $V_{biasRAD}$ til akkurat dette punktet, går man frem på følgende måte:

Først tildekkes kretsen, slik at det ikke kommer lys inn på fotodiodene. Dette gjør at alle fotodiodene ser et statisk (sort) bilde slik at alle endringspixelene trekker strøm. Vi setter så forspenningen for kolonnene til Vdd og lar den sakte gå nedover mens vi hele tiden følger med på kolonnespenningen V_{kol0} eller V_{kol1} (ikke forspenningen for kolonnestrømmen). Denne spenningen vil hele tiden ligge lavt fordi hele strømmen i kolonnen vil forsvinne inn i endringspixelene. Etter hvert som vi senker forspenningen og dermed øker kolonnestrømmen vil vi nå et nivå der V_{kol0} (V_{kol1})

stiger veldig. Vi har da et lite overskudd av strøm som gjør at nevronene sender ut pulser. Fra det punktet settes spenningen litt tilbake, i størrelsesorden 50 – 100 mV.

Deretter går vi frem på akkurat samme måte for radene. Om man stiller inn radene eller kolonnene først er likegyldig. Merk at det vil alltid komme noen pulser fra nevronene selv om rad- og kolonnestrømmene er mindre enn det endringspikselene ønsker å trekke p.g.a. parallellkoplingen. Dette er forklart tidligere under ”inaktivitetsnevronet”.

4. $V_{inactive}$ er den siste forspenningen som må settes. Under innstillingen av de tre første forspenningene kan de lønne seg å la $V_{inactive}$ være satt til Vdd. Effekten av dette er at inaktivitetsnevronene er helt koplet ut, siden de ikke får noen strøm inn, og vil ikke forstyrre innstillingen av rad- og kolonnestrømmene i punkt 3.

Kretsen er fremdeles tildekket fra punkt 3, så det kommer ikke noe lys inn på kretsen. Det vil likevel være aktivitet blant nevronene, i form av uønskede pulser. Spenningen $V_{inactive}$ senkes gradvis til inaktivitetsnevronet vinner i ”winner-take-all” kretsen, og dermed fjerner de uønskede pulsene. Kretsen trenger ikke lenger tildekkes.

10.4 A time-domain winner-take-all circuit of integrate-and-fire neurons

Winner-take-all kretsen som ble utviklet sammen med kameraet er innsendt til konferansen 2004 IEEE International Symposium on Circuits and Systems, som finner sted i Vancouver, Canada 23-26 mai 2004.



M:\hovedfag\
publication\paper_fin

